

DOCUMENT RESUME

ED 237 106

IR 050 527

AUTHOR Dietschmann, Hans, Ed.
TITLE Representation and Exchange of Knowledge as a Basis of Information Processes. Proceedings of the International Research Forum in Information Science (5th, Heidelberg, West Germany, September 5-7, 1983).
INSTITUTION German Society for Documentation, Frankfurt (West Germany).
PUB DATE Sep 83
NOTE 418p.
PUB TYPE Collected Works - Conference Proceedings (021) -- Reports - Descriptive (141) -- Viewpoints (120)
EDRS PRICE MF01/PC17 Plus Postage.
DESCRIPTORS *Artificial Intelligence; Automatic Indexing; *Cognitive Processes; Computer Oriented Programs; Foreign Countries; *Information Needs; *Information Retrieval; *Information Science; Library Research; Management Information Systems; Man Machine Systems; *Online Systems; Technological Advancement
IDENTIFIERS *Knowledge Representation

ABSTRACT

This 22-paper collection addresses a variety of issues related to representation and transfer of knowledge. Individual papers include an explanation of the usefulness of general scientific models versus case-specific approaches and a discussion of different empirical approaches to the general problem of knowledge representation for information systems. Specific topics include advanced analysis procedures for automatic indexing and abstracting, the interdependencies of communication to the structure of language and actions, the influences of technological impacts on information needs and the treatment of social problems, the intellectual architecture of information systems, designing "crutches" for communication, text-understanding as a knowledge-based process, heuristic text parsing in knowledge-based text understanding and condensation systems, an information flow model as the basis for a workstation for the problem solver, online machine interaction facilities, the Education Management Information Exchange, using MONSTRAT (modular functions based on natural information processes for strategic problem treatment) in the analysis of information interaction, two-dimensional representation of information retrieval services, and design of a self-organizing documentation system. The use of information systems in planning, administration, and management is discussed in the context of case studies. Papers were presented by participants from Australia, Finland, West Germany, Italy, the Netherlands, Switzerland, the United Kingdom, and the United States. (LMM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PREPRINTS

IRFIS 5

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

H. Dietschmann

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

FIFTH INTERNATIONAL RESEARCH FORUM
IN INFORMATION SCIENCE

SEPTEMBER 5 - 7, 1983

"Representation and Exchange of Knowledge
as a Basis of Information Processes"

DEUTSCHE GESELLSCHAFT FÜR DOKUMENTATION E.V. (DGD)
Sektion Informationswissenschaft (SIW)

in cooperation with

GESELLSCHAFT FÜR INFORMATION UND DOKUMENTATION
MBH (GID)

TAGUNGSHOTEL MOLKENKUR
IN HEIDELBERG, FEDERAL REPUBLIC OF GERMANY

FRANKFURT

2

SEPTEMBER 1983

ED237106

IR050527

Editor:

Dr. Hans Dietschmann

Coordinator, IRFIS 5

Gesellschaft für Information und Dokumentation

Lyoner Straße 44-48

D-6000 Frankfurt/Main 71

Federal Republic of Germany

Contents

Programme

G. WERSIG, R.-D. HENNINGS (Freie Universität, Berlin)

The intellectual architecture of information systems. A broad-range research agenda.

L.C. SMITH, A.J. WARNER
(University of Illinois, Urbana-Champaign)

W. KUNZ (Gesellschaft für Information und Dokumentation, Heidelberg)
H.W.J. RITTEL (Universität Stuttgart, University of Berkeley)

How to know what is known: Designing crutches for communication.

M. KALLFASS, TH. SEEGER (Freie Universität, Berlin)

The stereotype approach of knowledge representation.

B.C. BROOKES (London)

The relevance of Laplace's "rule of succession" to knowledge representation.

R. KUHLER (Universität Konstanz)

Some similarities and differences in intellectual and machine text understanding for the purpose of abstracting.

C.R. ROLLINGER, H.-J. SCHNEIDER (Technische Universität, Berlin)

Text understanding as a knowledge-based process.

U. HAHN, U. REIMER (Universität Konstanz)

Heuristic text parsing in knowledge-based text understanding and condensation systems.

R. ADAM (The City University, London)

The social construction of sociological language and its implications for communication.

(Paper not submitted at the conference)

M. BRITTAIN (Loughborough University)

Penalties for ignorance in the medical sciences and health care implications for information transfer.

R.T. BOTTLE (The City University, London)

Declining readability as a factor limiting the transfer and use of knowledge.

M. BUDER, G. VOWE (Freie Universität, Berlin)

Treatment of societal problems and information systems. An expanded concept of "Information- and communication technology".

K. JÄRVELIN, A.J. REPO, P. VAKKARI (University of Tampere)

On the implications of modern information technology to information needs.

A.R. BLICK (Beecham Pharmaceuticals, Betchworth)

Research in information science versus the practitioner

F.C.A. EXON

(Western Australian Institute of Technology, South Bentley)

Methodological problems in user studies, and the organisation of in-house library research.

T.D. WILSON (University of Sheffield)

Environmental monitoring and information management in local authority planning departments.

D. STREATFIELD

(National Foundation for Educational Research, Slough)

The Education Management Information Exchange (EMIE): A user-based research project aimed at local authority education administrators.

R. McCALL JR., B. SCHAAB, W. SCHULER
(Gesellschaft für Information und Dokumentation, Heidelberg)

An information flow model as basis for a workstation for the
problem solver.

P. INGWERSEN (ESA, ESRI, Frascati)

Online man-machine interaction facilities: a cognitive view.

N.J. BELKIN (The City University, London)
G. WINDEL (Freie Universität, Berlin)

Using MONSTRAT for the analysis of information interaction.

H.P. FREI (ETH Zurich)

Two dimensional representation of information retrieval
services.

J.L. HOROWITZ (Rutgers University, New Brunswick)

New technology, scientific information, and choices for demo-
cratic societies.

G.J. DALENOORT (University of Groningen)
Design of a self-organizing documentation
system.

(This contribution was integrated instead
of the contribution of G. LUSTIG)

THE INTELLECTUAL ARCHITECTURE OF INFORMATION SYSTEMS
A Broad Range Research Agenda

G. Wersig
R.-D. Hennings
Projekt INSTRAT
Free University, Berlin

1. Introduction

We wish to introduce a way of considering information systems which we think is likely to be productive in both design and research terms.

In the first part of this paper the design criteria for information systems are presented. Here we refer heavily on the concept of a "problem treatment process".

Then we describe the MONSTRAT-approach and some problems in the context of designing and implementing some components in INTERLISP. In the last part a broad research agenda is presented.

This approach to information systems arose in the context of Projekt INSTRAT, a research project funded by the Federal German Ministry for Research and Technology. Projekt INSTRAT is concerned with "information systems as information policy shaping potential and social development strategy - the information science bases of organized information and communication as components of individual and social problem management". This article is partly based on reports from that project (1,2,3).

2. Design criteria for information systems

The traditional view of "information systems" concentrates quite often on technological devices to receive, store and process information in order to provide it for someone. This view is strongly influenced by computer science and its applications.

Professionals would usually expand this view toward the institution which is holding the technology, serving it, and offering it to its "users". Nevertheless, this is a very limited view.

The starting point for all our considerations is the situation of humans

having problems. For the treatment of these problems they might need help of information providing mechanisms (IP).

In this context the following considerations are essential :

- There is no specific "information need" but there are people in problematic situations where at different states of the problem treatment process different kinds of information provision may be helpful.
- The individual view of the "problematic situation" is very often biased, incomplete, and limited. Thus information provision can help in structuring the problems view.
- The situations, where information provision may help the individuals in its problem management are not restricted to simple reference, document or fact provision, but may require advice, counseling, proof, identity assurance etc.

That requires a new approach to information provision with the following features :

- Since problem treatment processes are dynamic, information provision has to be feasible of adapting itself to its changing states. This feature could be called **diachronic**.
- As different people have distinct problems, the information provision has to adapt itself to the levels required in the respective cases. This feature could be called **telescopic**.
- Information provision itself is not neutral to the problem treatment process. Therefore it may take functions of helping the user to change their view of the problem. This feature could be called **maieutic**.
- Finally information provision under these conditions is a process where people in need and the information provision mechanism or agency focus on each other, which could be called the **interactive feature**.

Existing so-called information systems mainly fulfill only some of these features: Reference retrieval systems for instance assume a specific, very restricted need situation, they do not react specific to different users, their interactive capacity is restricted to a formal level etc. Thus, in interaction with traditional systems like this all of the functions which are really important in a problem treatment process of people in need are burdened upon them.

Until now only humans are capable of doing the above mentioned information provision. But information provision by humans is also not feasible for a growing number of tasks in everyday life of highly industrialized and civilized countries. The reasons for this are numerous, e.g. growing amount of knowledge produced, growing quantity and quality of different knowledge needed for everyday and working life, increasing complexity of communication channels, increasing spatial dispersity of persons, who need each other in information processes.

3. Problem treatment process

This, to some extent, has been caused by technologies and scientific results and the ways they are used. So it seems to be fair of thinking how technology and science could be used to reduce the above mentioned effects in future. The traditional approach was to think of one function or one problem and then to try to "solve" it.

We prefer to use the term problem "treatment", because there seems to be a large class of information-related situations in which appropriate action is not simple problem "solution", but, for instance, explication, re-definition, denial, and so on.

The implications of this approach are strong, and we would like briefly to outline those that have been important to us.

In particular, we notice that in the problem treatment framework, it is as important for the information mechanism to know about:

- problem resolution strategies,
- problem dimensions,
- interaction modes and strategies,
- problem solution alternatives,

effects of decisions,
aims of the user,
feedback mechanisms,

as it is for it to represent and deal with factual or reference knowledge.

Being able to provide the user (and to utilize) these kinds of information allows the mechanism partially to fill the gap between knowledge resources and problem resolution. In particular, the mechanism can be helpful in assisting the user to progress from an ill-formed problem to a well-formed problem (4).

Thus, the problem treatment approach implies that:

- the information mechanism interacts **iteratively** with the user, assisting her/him toward appropriate treatment of the problem by providing information which is appropriate in terms of the individual solution path and the characteristics of the problem situation; and,
- the information mechanism is **multi-functional**, being concerned at least with problem definition and redefinition, aim definition, interaction modes and resolution strategies, as well as with representation and retrieval of factual or reference type knowledge.

We note that empirical research demonstrates that these functions actually appear in information interaction, (5) and suggest that they should be considered explicitly in the design and analysis of any information system, whether the functions are fulfilled by humans or by machines.

4. The MONSTRAT-Approach

The problem of a general model of information systems which uses basic concepts in a way that in reality they could be represented by human or technological components has been led to the MONSTRAT-approach.

Experiences and research with information seeking people, intermediaries, and interaction between both, have revealed some basic functions which have to be fulfilled by any information provision mechanism.

The modelling of these functions followed approaches of artificial intelligence, where systems are constructed out of separate **experts**, which are responsible for a specific task. The communication among these experts can be realized by different models.

This led to a model of "Modular functions based on natural information processes for strategic problem treatment (MONSTRAT)".

The MONSTRAT-model of the intellectual architecture of the information provision mechanism within an information system is thought of an interactive process between an information seeking (IS), normally a user, and an information providing (IP) mechanism being realized by modular components, each of which might be represented by

- computerized or computer-aided function(s) within a main frame unit or on distributed systems
- human function(s) by a single dedicated person
- social function(s) being dealt within a social mono- or heterofunctional group.

Thus, the basic modules can be specified in the following manner : which is grafically represented in figure 1 .

As far as a implementation is considered the concept of "experts" is used synonymous to "function" and "module".

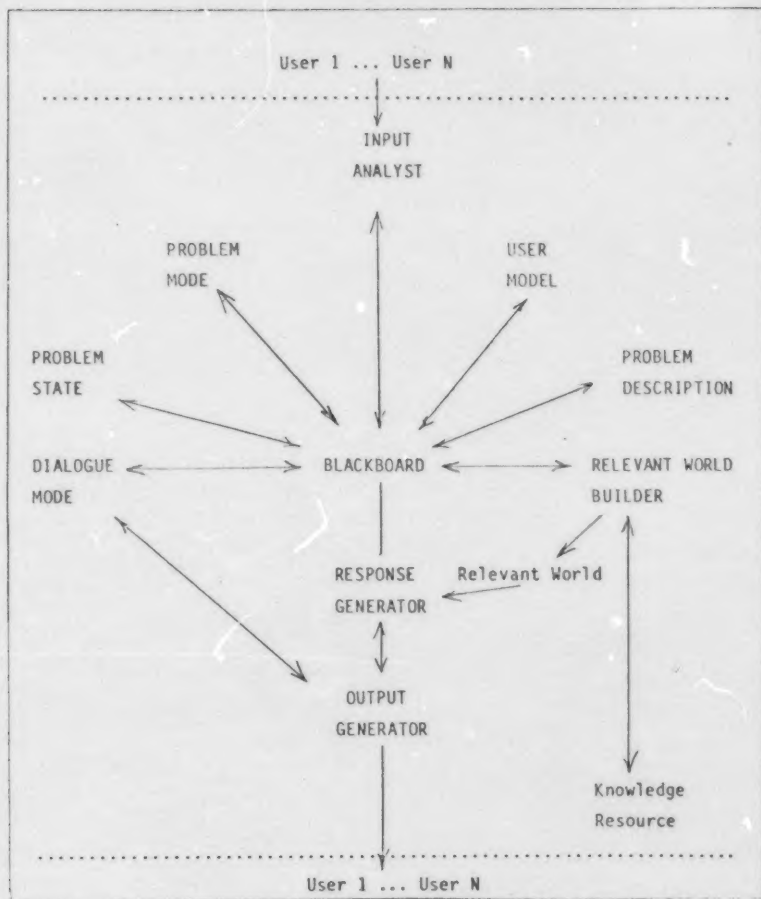


figure 1: MONSTRAT-Model

4.1 Input function

Regardless of how the interaction should be started (the recent model expects the IS to start, but this is optional) interaction requires input from the IS-mechanism into the IP-mechanism.

To be processed and "understood" internally by the IP-mechanism this input has to be transformed in an appropriate internal representation and then analysed according to certain aspects like.

- language and language-level being used
- dialogue mode and
- linguistic skills being indicated.

This is the task of the **input analyst (IA)**.

4.2 Analytical functions

The analytical functions have to analyse the transformed input according to different tasks. Investigation of "natural information processes" so far revealed the following functions which have been verified by a simulation with human experts.

4.2.1 Problem description

The **problem description (PD)** function of the information mechanism is concerned with describing or modelling the user's problematic situation (6) or anomalous state of knowledge (7) in such a way that it is capable of discovering the type of problematic state or anomaly which has brought the user to the mechanism, and so that the mechanism can identify the topic(s) of the problem and its context.

The topical function need not be discussed here, as it has been extensively considered in the literature: by type of problem we mean, for instance, procedural or decision making or learning; by structure we mean coherent or unstructured state of knowledge, gap in knowledge or need for establishing relations, and so on. This categorization is necessary because different problem types might require different strategies for retrieval from the mechanism's data base.

4.2.2 Problem state

Another function must be to understand where in the problem-treatment process the user is located. For instance, has the user already defined the problem s/he intends to treat? is the problem **ill-formed** or **well-specified**? and so on.

This function is necessary because users at different points in the problem treatment process will require different kinds of information, and different kinds of responses from the mechanism. For example, a person just beginning to understand a problem might require help in shaping the problem itself, whereas a person with a well-defined problem might require only a factual response. We call this the **problem state** function (PS).

4.2.3 Problem mode

Related to the problem state function is the requirement that the mechanism be capable of choosing the appropriate mode of response: that is, of knowing what it is capable of doing, and of choosing the appropriate capability for the given situation. Thus, capabilities of a general information provision mechanism might be fact retrieval, problem specification, advice or comparative evaluation, role reinforcement, and so on. An important aspect of this function, the **problem mode** function (PM), is that the mechanism be able to recognize situations for which it is not appropriate for which it has no appropriate capabilities - and that it be able to respond accordingly.

4.2.4 User model

The state and mode functions are closely related to the user's problem-treatment activity and to how that activity relates to the mechanism's problem-treatment capabilities. In addition to this type of knowledge, there is a requirement for knowledge of specifics of the situation, related to details of the user and the problem the user is facing. If we consider the interaction between mechanism and user as a cognitive communication system, as does Hollnagel (8), then we see that one of the primary functions of the mechanism must be to develop a model of the user, especially of the user's intentions, situation, preferences and, perhaps, beliefs. That is, a model of what the user is like, and therefore, of what the user is likely to need.

For instance, if it is known that the user is a beginning research student, this will be of great help to the mechanism in choosing an appropriate response. The response in such a case is likely to be quite different than the type of response, even in the same general subject area, that would be appropriate for an experienced researcher or an interested lay-person. Thus, the user model function (UM) is concerned with developing a description of the user and the user's intentions.

These four functions constitute the more **analytically** working functions. The following modules are responsible for the generation of "solutions" - as far as they exist - of the problem treatment process, i.e. a response of the IP- to the IS-mechanism.

4.3 Synthetical functions

4.3.1 Relevant World

The response is based upon the mechanism's model of the world: that is, its knowledge or data base, and will depend upon the mechanism's perception of what aspects of its world are relevant to the particular user in the particular situation. Thus, a major function must be the interaction of the mechanism with the **knowledge resource** to decide upon a partition of the world relevant to this situation.

This function we call the **relevant world builder (RWB)**. Notice that the relevant world builder will have a set of strategies for interaction with the data base, the choice of strategy being dependent upon knowledge of the whole situation. Again, if there appears to be no part of its world relevant to the situation, the mechanism must be able to inform the user, and to suggest alternative strategies or treatment modes.

4.3.2 Response generation

It is insufficient for the purposes of the mechanism, therefore, for there to be only a representation of the **relevant world** to convey to the user. It is necessary to interpret that relevant world in terms of what the mechanism knows of the user and the user's situation, so that the mechanism can generate an appropriate response. For instance, if a user asks for the time of

the next train to another city, and the relevant world indicates that the last train of the day has already left, then it might, in some circumstances, be appropriate for the mechanism to offer additional information about other forms of public transport, or even just how to get in contact with persons in the other city, as well. This function, depending to a great extent upon information provided by the other functions, we call the **response generator (RG)**.

4.3.3 Dialogue mode

In order for the mechanism to accomplish the previous functions, it is necessary for it to interact appropriately with the user. This will be accomplished via some form of dialogue, but the specific kind of interaction used in any context will depend upon the circumstances of the situation. Dialogue is used to elicit information from the user, which is then used by the information mechanism to help it accomplish its functions; dialogue is used to help the user and the mechanism in problem formulation; and, dialogue is used to inform the user and the mechanism of one another's progress, intentions, perceptions of the other and of the problem, and so on.

In general, dialogue is used to help both the mechanism and the user to manage the user's problem effectively. For different types of users in various problem states and with various types of problems, different modes of dialogue will be appropriate (indeed, different modes will be appropriate for the mechanism according to its state).

For instance, at some states a natural-language interaction in an informal mode might be appropriate (in, say, initial problem description by the user with as yet nebulous perceptions of the problem), in others, interactive graphic interaction might be preferred (in, say, representation of the mechanism's perception of the user's anomalous state of knowledge for a user who prefers multi-dimensional to linear representations). The **dialogue mode** function (**DM**) is to determine the type of dialogue appropriate to the given context.

4.4 Output

Last but not least the **output generation** function (OG) converts the structures which the other functions have achieved into some form which the user can understand, in a way that is appropriate to the immediate situation (in response to the dialogue function).

4.5 Auxiliary functions

One service function is the **explanation** (EX) in which the information mechanism informs the user of its capabilities, restrictions, method of operation, and so on.

As far as the communication mode uses a blackboard, a second function is necessary: the **blackboard analyst** (BA). This module is responsible for a well ordered blackboard.

Another function is responsible for the various administrative or logistic details appropriate to the situation, such as cost of the service, etc. It is called **system** (SYS).

We believe that the functions we have enumerated (see figure 1):

input analyst	(IA)
problem description	(PD)
problem mode	(PM)
problem state	(PS)
user model	(UM)
dialogue mode	(DM)
relevant world builder	(RWB)
response generator	(RG)
output generator	(OG)

and three auxiliary functions which are not represented graphically

explanation,	(EX)
blackboard analyst	(BA)
system	(SYS)

constitute the minimum functions required of a general information mechanism, if it is to carry out its task of helping a user to manage problems appropriately. As justification of this view, we suggest considering the interactions and other activities inherent in any good reference interview or advising interaction.

Our functional analysis suggests that the information mechanism might be considered as a number of interacting but discrete functional components. Notice, however, that these components interact with one another in highly complex ways, each depending upon information from all, or almost all the others, at almost all stages of accomplishment of the individual functions.

5. Design and Implementation of the MONSTRAT-System

Though we have stated above that all of the functions of the IP-mechanism, which we have isolated, can only be fulfilled by humans until now, we have tried to design some elements, i.e. control structure and interaction mechanism to get at least insight in problems on this level.

On the other hand we can base our work on results from other research groups, which are mainly involved in studying the functions of some experts, for instance the input analyst or the output generator, with greater amount of details.

Nevertheless it seems important to us to transform part of the ideas, presented in the preceding sections of this paper on a constructive level, thus revealing problems, which force us to seek for solutions as far as it is possible.

During this work we are mainly influenced by ideas, concepts and methods coming out artificial intelligence. Especially the results from different efforts to develop **expert systems** are of great interest for our research. Here we have studied the underlying concepts of the different frameworks of distributed problem solving (9) and other ideas (10). As consequence we consider all above mentioned functions being realized by different independent working **experts**

5.1 Network topology

In a first step we have investigated a **network** for the MONSTRAT-system being appropriate for both the communication types we have simulated: the blackboard-model and the actor-model, both being described below.

In this first version of the modelled MONSTRAT-system the nodes of the network represent the different experts whereas the arrows describe the relations between the experts as far as communication is concerned.

Two different communication modes are considered:

The **blackboard-model** was introduced in (11). We have adapted the underlying ideas to our special functions of the problem treatment process: in one communication mode all experts are permitted to communicate only via a central blackboard, which is synonymous for a certain amount of memory space to hold the respective output of the different experts. This communication can in a certain sense be called an **indirect** one.

On the other hand we have used the concepts of the **actor-model**, being introduced by (12). Here all the communication is done expert-to-expert: thus we call it a **direct** one.

Some insights concerning the communication modes were derived from a simulation of the whole system by humans and were presented in (13,14,15,16,17).

Very important ideas for the MONSTRAT-system were introduced through the architecture of **Open Systems Interconnections**. Especially the concept of "open" systems seemed to be important for us. Therefore we developed a

layered structure being extendable at choice such that the number of functions is not limited.

Another point to mention in this context is that the system should be a distributed one. Here some difficulties arise because there are some unsolved problems concerning the synchronization and cooperation of "tasks" or "processes", when they are **independent** from one another what means that there is no master-slave-relation between them (18).

5.2 Control structures

On the basis of the network we developed **control structures** being appropriate for different levels of the model. As far as an implementation of parts of the model is planned, we assumed to use several computers of the fourth generation or - when being available one of the fifth generation.

So we have to introduce one or more levels of control between the machine and the system. Additionally at least one level of control must be placed between the system's level and the respective experts.

The flow of control was described in two different models: A model for the **static** behaviour uses concepts of automata theory, i.e. states, transitions, input- and output-signals. For representing the **dynamic** properties of the system we used **procedure-graphs** (19), which are a special variant of "free choice nets" (20) respectively a subset of Petri-nets (21).

Petri-nets and procedure-graphs make use of two different states: one for **information-items** and one for **actions** being represented graphically either by circles or by squares.

Problems arise when the procedure-graph should delineate cyclic processes, which is excluded in the models being developed in (22).

5.3 Interaction

Especially when **interaction** between the different experts is concerned we base our ideas for an implementation on material from the above mentioned simulation.

All communication was realized by means of written messages in order to be able to evaluate the different interactions afterwards.

In a first approximation we isolated six different **types of messages** i.e. questions, answers, hypotheses, context information, knowledge in form of facts and rules.

These types could initiate different **sequences of actions or reactions**, being reduceable to five different patterns, which could be arbitrarily superposed.

On the basis of these patterns we tried to represent the different interactions which were recorded during the five rounds of simulation.

Especially two types of interactions occurred :

either an expert-to-expert interaction, for example question and answer, or the case when the action of N experts permits the reaction of M other experts.

By means of several constructive elements on a software-level we are able to reproduce the interactions having taken place during the simulation by humans.

5.4 Implementation with INTERLISP

As far as pointed out in the preceeding sections of chapter 5 we are going to implement different modules/functions in INTERLISP on a SIEMENS 7.541. Just now we are specifying single functions using the proposed frame structure being explained in (23).

5.5 Implementation outlook

Since distributed expert systems, related communication and control structures are a similar approach to several specific problems usually considered to be an object of artificial intelligence, it could be expected that general principles be developed, tested, and implemented for several purposes other than the MONSTRAT approach.

Specific problems which may be particular of interest for information provision mechanisms are e.g.

- levels of control
- direct vs. indirect communication
- message types and their treatment in the different communication modes
- parallel vs. sequential actions of the experts.

They may be implemented soon.

6. Open brainware questions

Information and communication technologies will in not too much time provide us with hardware facilities for information intelligence and communication comfort which even were not dreamable for an information professional at the time when they started to establish forerunners of mechanized information systems. These facilities only can be used appropriately if we start early enough to develop the respective brainware foundations. The MONSTRAT-approach is a first step in this direction outlining some basic features of the intellectual architecture of information provision mechanisms which become thinkable from now on.

To follow these approaches in the line of construction a lot of disciplines are challenged to provide the necessary input: information science, artificial intelligence, cognitive science, only to name a few. The challenge of research and development by a concept like MONSTRAT could obviously not be met by a single research group but requires work on foundations and components at various places, using every experience and expertise that is available.

In fact, this would require a broad range research and development programme being executed by an instance on a high research policy level and on the

highest level of competence. The ESPRIT-programme could have been such a chance but was wasted to some extent by the obvious lack of any integrative aim and theory. The only hope we have is that by and by the respective research community itself will see the advantages of a concept which permits to integrate the bits and pieces which factually constitute the research and development work in this area.

The basic assumption behind the MONSTRAT-approach is that technology in the future would permit the modelling of information provision mechanisms not as features of a rather dull technology but requires that we have to know much more about humans in problem treatment situations than we already know in order to develop the respective models and processes which then could be implemented. The MONSTRAT-approach in this context could be used as a generator of a respective research agenda.

Some of the problems which instantly come to the mind of the agenda builder will be mentioned in the following, this being by no means a comprehensive and final list.

6.1 User models

Modelling of users so far has concentrated on some social characteristics, since its use was restricted (if they ever were used) for purposes of marketing or reasoning with funding agencies. A MONSTRAT-mechanism would need user models which concentrate on other, non-trivial features which relate to the users problem-treatment and interaction behaviour.

Some elements of user models of this type may be

- the "cognitive styles" of users, that means their individual approach of thinking, reasoning, discrepancy reduction
- the style of problem treatment and problem solving of users, like decisionist, cunctator, autocratic and democratic problem solvers, etc.
- the different forms and intensities of motivating forces within users.

In all cases, and certainly all of You could add easily more items to this list, we have to see that in most cases not only one kind of model is possible but several alternative approaches of model building will be necessary,

to name only some like ideal types, structural models, procedural models, layer models.

In addition to that, we do not need only to develop these models but need to define sets of indicators in interaction which allow to analyse messages in order to develop hypothesis which model may be applicable, we need to develop these hypothesis forming operations, we need operations to develop questions put forward to the interactive partner to gain information about him which is essential to the hypothesis formulation (and we need questions to find out what is essential).

On the other side we need rules which permit us to use these models and their potentials in the process of building relevant worlds and generating answers.

To make the problem more complex I have to remind You that, in addition to that, our mechanism of course needs in the area of user models a memory of users he already knows, which has to be updateable and changeable, he needs a learning mechanism to evaluate the success of its operations, improve his analytical and evaluative potentials and procedures, and to refine its models.

This seems to be a rather complex research agenda, although perhaps user models itself seems to be the most simple and least crucial part of the whole agenda, since we already know a lot about users and at least via the intermediaries have a lot of experiences available which could be used. The following topics are fewer and therefore at that time less to be structured, so that I only outline some basic questions just to indicate the direction.

6.2 Problem treatment processes (PTP)

We hardly know how people treat problems in reality with respect to information needs being generated or suppressed in this process. The model we developed is an unempirical one, what we now need is to go into empirical work and study PTP's in order to find out

- different approaches to PTP, PTP-modes or styles
- stages of PTP, sequences, possible loops, by-ways, short cuts etc.
- the role of different kinds of information in different approaches and stages of PTP's

- the role of driving motivational forces in PTP's, their relation to information need formation or suppression.
- components of PTP's like hypotheses, cost-benefit evaluations, aims, resource analysis etc. and their relation to different kinds of information.

The basic problem in my view would be to give up the illusion of economics that the problem carrier is a rational being ("economic man" or "information man"). Problem carriers have to be taken as what they are - information provision cannot rely on rational behaviour because it itself is a step towards rationalization of action.

6.3 Problems

The very basic concept underlying information science is not "information" but "problem". In fact, we do not know very much about problems, so that a very important need that has to be satisfied would be to know about

- the structure of problems (it seems that the assumption of a dilemmatic structure could be very helpful)
- types of problems
- ways of representation of problems
- methods to develop problem descriptions
- maieutics of problem structuring.

Although more clarity about the nature of problems itself is a problem of high complexity it only would be the very basis of the essential question whether it could become possible to establish relations between problem structures and preferable problem treatment strategies to be applied by the information provision mechanism.

6.4 Information representation

Obviously the information provision mechanism I am talking about could not function very satisfactorily with the forms of knowledge representation we are used to in present systems. The information basis of such a mechanism would have to cover

- very different forms of information sources (from texts to sounds)
- very different forms of information representation (from citations to data)
- very different forms of information (from knowledge indicators to values and esthetics)
- very different forms of knowledge structuring (from descriptors to graphics).

Ideally, this information basis is learning from its use and user, would be updated automatically and is flexible enough that any operations could be applied to it.

6.5 Construction of relevant worlds

The essential switching centre of the mechanism would be the combination of the different hypotheses of the analytical experts and using this integrated model of the current case to create out of the general information base a relevant world for this case. Unless we know a little bit more about these two areas it does not seem possible to formulate assumptions how this might be solved. In general, it might in itself be structured as a multi-faceted, multistep problem-treatment process.

6.6 Generation of answers out of the relevant world

The final step of the information provision process would be to operate on the relevant world in order to generate an answer to the statement brought forward by the user. A lot of experience already has been gained in artificial intelligence as far as classical logical operations are concerned. What is needed is the study and modelling of non-classical logics which are an

essential part of human cognitive operations. By "non-classical logics" we mean operations like heuristics, analogy, metaphor, even perhaps intuition (or some equivalent of it).

6.7 Interaction strategies

Finally, I would like to point to the area of interaction itself where it would be necessary to develop

- integrated interaction modes
- interaction strategies for different stages of the interaction process
- interaction strategies with reference to user types, problem types, etc.

Only to indicate how selective this list is, I have to draw our attention to some areas which I did not deal with, although they are part of the brainware research agenda like

- language, speech, pattern recognition, analysis and synthesis which would be needed for input analysis and output generation
- internal control and diagnostic procedures
- generation of the system behaviour explanation
- internal language of the system.

Although my remarks here are far from being comprehensive they already indicate the broad range of research and development that becomes visible by the MONSTRAT-approach. It offers to us a fascinating agenda with enough topics for everyone and for a long time. We only have to pick it up.

References

1. Wersig, Gernot: "MONSTRAT: The Intellectual Architecture of Information Systems", in: R. Trapp (ed.), Cybernetics and System Research, Amsterdam et.al. 1982, pp. 837-841.
2. Seeger, Thomas (ed.): Information Systems as Collectives of Cooperating Functional Experts, Proceedings of a Workshop, Berlin, 1981. FUB-IFP 5/82 Freie Universität Berlin, Fachbereich Kommunikationswissenschaft, Arbeitsbereich Informationswissenschaft, Berlin, 1982.

3. Belkin, N.J., Seeger, T., Wersig, G.: "Distributed expert problem treatment as a model for information system analysis and design", in: Journal of Information Science, 5, 1983, pp. 153-167.
4. Simon, H.A., "The structure of ill-structured problems", in: Artificial Intelligence, 4, 1973, pp. 181-201.
5. Brooks, H.M., Belkin, N.J.: "Using Discourse Analysis for the Design of Information Retrieval Interaction Mechanisms", in: Proceedings of the Sixth Annual International ACM SIGIR Conference, New York, 1983.
6. Wersig, Gernot: "The problematic situation as a basic concept of information science in the framework of the social sciences - a reply to N.Belkin, in: New Trends in Informatics and its Terminology. FID 568, VINITI, Moscow, 1979, pp. 48-57.
7. Belkin, N.J.: "Anomalous states of knowledge as a basis for information retrieval", in: Canadian Journal of Information Science, 5, 1980, pp. 133-143.
8. Hollnagel, E.: "The relation between intention, meaning and action, in: MacCafferty, M.; Gray, K. (eds.), The Analysis of Meaning: Informatics 5, Aslib, London, 1979, pp. 135-147.
9. Smith, Reid G.: A Framework for Distributed Problem Solving, Ann Arbor, 1981.
10. Stefik, Mark et.al.: "The Organization of Expert Systems, A Tutorial", in: Artificial Intelligence 18, 1982, pp.135-173.
11. Erman, L.D.; Hayes-Roth, F.; Lesser, V.R.; Reddy, D.R.: "The HEARSAY II speech understanding system: Integrating knowledge to resolve uncertainty", in: ACM Computing Surveys, 12/2, June 1980.
12. Hewitt, Carl: "Control Structure as Patterns of Passing Messages" in: Winston, P.H.; Brown, R.H. (eds.), Artificial Intelligence: an MIT Perspective Vol 2, Cambridge et.al., pp. 433-465.
13. Seeger, Thomas et.al.: MONSTRAT-Simulation I. Das operationalisierte MONSTRAT-Konzept. Projekt INSTRAT, April 1983 (A 1/83).
14. Seeger, Thomas; Hennings, Ralf-Dirk: MONSTRAT-Simulation II. Die Anlage der Simulation. Projekt INSTRAT, April 1983 (A 2/83).
15. Hennings, Ralf-Dirk et.al.: MONSTRAT-Simulation III. Register und Materialien. Projekt INSTRAT, Juni 1983 (B 3/83).
16. Hennings, Ralf-Dirk: MONSTRAT-Simulation IV. Übergreifende Auswertung. Projekt INSTRAT, April 1983 (A 4/83).
17. Belkin, Nick J.: MONSTRAT-Simulation V. Statistical Evaluation of the MONSTRAT Simulation. Projekt INSTRAT, April 1983 (A 5/83).

18. Schindler, S.; Burkhardt, H.J.: "Konzeptionelle Grundlagen strukturierter verteilter Applikationen.", in: Schindler, S.; Spaniol, O.: Kommunikation in verteilten Systemen - Anwendungen und Betrieb. GI:NTG-Fachtagung, Berlin, Januar 1983.
19. Kreifelts, T.: "Coordination Procedures: A Model for Cooperative Office Processes, in: Schindler, S.; Spaniol, O.: see ref. 18.
20. Hack, M.H.T.: Analysis of Production Schemata by Petri Nets. MIT Project MAC TR 94, Cambridge/Mass. 1972.
21. Petri, C.A.: Kommunikation mit Automaten. Schriften des Institutes für Instrumentelle Mathematik, Bonn, 1962.
22. Kreifelts, Thomas: "Coordination procedures", IIG-Report, Gesellschaft für Mathematik und Datenverarbeitung, St. Augustin, 1982.
23. Hennings, Ralf-Dirk: MONSTRAT-Designbericht 3. Grundlegende Ansätze der Implementierung von MONSTRAT-1 mit INTERLISP. Projekt INSTRAT, Juni 1983 (B 3/83).

A Taxonomy of Representations in Information Retrieval System Design

Linda C. Smith and Amy J. Warner

University of Illinois at Urbana-Champaign
Urbana, IL 61801/USA

1. INTRODUCTION

This paper is a preliminary report of a study which has had the objective of developing a taxonomy of representations in information retrieval (IR) system design. In IR there is a growing body of empirical data and practical experience with various representations. Our study involved a review of the principal English language journal and conference literature dealing with IR system design and experimental IR. Analysis of this literature included identification of what was represented, how it was represented, and the results of any evaluations. A comprehensive final report of the study is in preparation. The purpose of this paper is to highlight major findings and to suggest areas in need of further research. Sections which follow include discussion of: the concept of representation in IR, the rationale for creation of a taxonomy of representations, major categories of representations, additional needed categories, and approaches to evaluation of alternative representations. The emphasis is on document (bibliographic) retrieval systems, because that is the focus of much of the literature reviewed.

2. REPRESENTATIONS

...a deeper understanding of how representations are created and how they contribute to the solution of problems will become an essential component in the future theory of design. (Simon, p. 153)

Before turning to consideration of the need for a taxonomy, it is necessary to clarify what is meant by "representation" in the context of IR system design. Simon and Siklóssy have observed that "representation has to do with

the kinds of symbol structures that are used inside a computer (or a human brain) to represent the external task situation confronting it." (p. 1)

Unlike linguists (e.g., Chomsky) or psychologists and philosophers (e.g., Fodor), our primary concern is not mental representations, but symbol structures used inside a computer.

As Cremmins notes, within the literature of information science, the word "representation" has frequently appeared as part of the phrase "document representation", but rarely has it been coordinated with "information" or "knowledge" to form "information representation" or "knowledge representation". The latter phrase, however, is now used extensively in the literature on artificial intelligence (AI), computational linguistics, and cognitive psychology. In fact, computer-based IR systems include representations of many things in addition to documents. Representations may include those for:

(1) objects; (2) relationships, which establish meaningful links from one object to another; and (3) processes, which govern the creation, destruction, transformation, and other behavior of objects (Raphael). All of these can be found in computer-based IR systems.

3. TAXONOMY

An early step toward understanding any set of phenomena is to learn what kinds of things there are in the set--to develop a taxonomy. This step has not yet been taken with respect to representations. (Simon, p. 154)

In providing the rationale for creating a taxonomy, it is interesting to begin by juxtaposing two statements regarding the state-of-the-art in AI and IR:

AI is still very much in the so-called "natural-history" stages of scientific activity in which specimens are collected, examined, described, and shelved. At some later time a theory will be suggested that unifies many of the phenomena noted previously and will provide a framework for asking questions. (Buchanan, 1982, p. 284)

It is arguable that our current understanding of information processing is like that of sixteenth century herbalists: it embodies some observation and insight, but lacks detailed analysis and supporting theory. (Sparck Jones, p. 3)

Just as natural history has advanced with the development of taxonomies for

plants and animals, IR (and AI) could benefit from the development of a taxonomy of representations. Whereas biological taxonomies deal with natural phenomena, IR systems are the products of man's inventiveness. In developing a taxonomy, one must be concerned with both classification and nomenclature, where classification is the process of establishing and defining systematic groupings and nomenclature is the allocation of names to the groups so produced (Jeffrey). The latter activity is important because one is most prepared to identify as objects or relationships those things for which one's vocabulary happens to contain a word. The presence of such a word can focus one's thinking onto what then appears as a singular phenomenon (Kent).

A number of papers have appeared in the computer science literature proposing taxonomies for such things as data models (Kerschberg et al.), computer interconnection structures (Anderson and Jensen), office systems (Conrath), and information systems (McKenzie and Spiers). Each provides a rationale for the creation of a taxonomy in the domain under consideration. Anderson and Jensen have noted "the almost total lack of published information describing the rationale for various designs, or comparing the results achieved by various approaches." They suggest that "this condition exists because there has been no common context in which such discussion could take place, no set of design issues, no list of system characteristics to be traded off, and, in fact, not even a common nomenclature for system identification." (p. 197) Similarly, Conrath has stated: "If research on the specification and evaluation of office automation technology is to be anything more than isolated individual efforts, we need to have a common basis on which to collect data so that we can exchange findings. The formation of taxonomies is the first step in this process." (p. 207) A taxonomy should provide a means of classifying things of interest according to their characteristics (McKenzie and Spiers) and a framework within which alternatives can be evaluated (Kerschberg et al.). It is thus a means

of arrangement and interpretation.

In our study the method of developing a taxonomy of representations was primarily inductive, establishing classes based on IR system descriptions found in the literature. Any taxonomy derived in this way must be regarded as tentative. Other items most likely will need to be added, and there may exist more fundamental ways of classifying its members. But even though our taxonomy is incomplete, and perhaps superficial, it can serve two important purposes: (1) it permits one to relate new work to that of previous authors who have already contributed to the field and (2) it allows one to detect voids and suggest new fields of endeavor. The remainder of this paper identifies categories of representations and areas in need of further study.

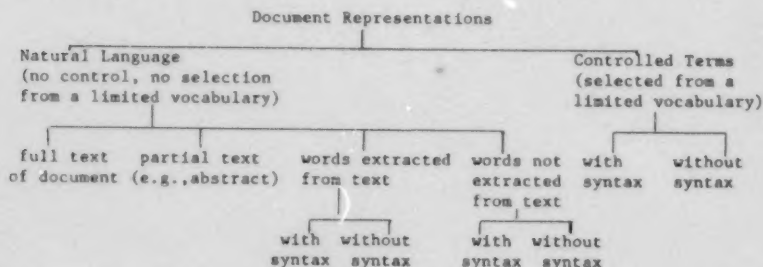
4. CATEGORIES OF REPRESENTATIONS

A computer system for IR is the result of a series of design decisions. The choices of representations are interdependent; clearly the form of query representation cannot be selected without regard for the form of document representation. Nevertheless, in this section each category of representation is discussed separately to present its salient characteristics. Although the categories described may all be part of an IR system, they may not all be found in any given system. Because the focus is symbol structures inside the computer rather than mental representations, no attempt has been made to characterize representations which may be used by the human indexer, abstracter, searcher, etc.

4.1 Document Representations

Because much of the literature reviewed dealt with document retrieval systems, it is possible to present a reasonably detailed list of alternatives for document representation. Until recently it has not been feasible to store the full text of many documents in machine-readable form, so there has been considerable interest in creating document surrogates as document representations.

Just as a cataloger carries out both descriptive and subject cataloging, most document representations contain both descriptive (e.g., author, publisher, date) and subject (e.g., subject headings, classification numbers) elements. The form of descriptive elements in the computer does not differ much from the form found in printed sources. Much more attention has been focused on subject elements, and a number of variations can be identified. The following taxonomic tree, taken from Lancaster, illustrates some of these:



The lowest nodes each represent a possible subcategory of document representations. The tree serves a useful purpose in identifying two ways in which document representations may differ (controlled/uncontrolled vocabulary, terms with/without syntax). Nevertheless, the subcategories presented are neither mutually exclusive nor exhaustive. Document representations may be hybrid including both natural language and controlled terms. A more complete tree would show further subdivisions, such as the use of codes vs. words as controlled terms.

The variety of document representations presented in the IR literature seems to indicate that novel approaches continue to be introduced. In reality there are few differences in the representation of documents as objects. The variety stems from the large number of different processes (algorithms) used to generate the representations through automatic indexing. In essence each process is a method for identifying what the designer deems to be "good"

indexing terms. Components of these processes may include such things as stop-word lists, stemming algorithms, and word frequency tabulators. Depending on the choice of the process for automatic indexing, rather different sets of terms (both in number and in membership) could be selected for inclusion in the document representation. The structure of the resulting document representations may be quite similar, however. The most common form is a list of words, phrases, and/or codes. Sometimes these are modified with weights to indicate relative importance. Sometimes context information is preserved, such as the field (e.g., title, abstract) from which the term has been extracted and even the position of the term within that field.

A final source of variations in document representations arises from the limited character sets which may exist. The available encoding for digits and roman characters must be supplemented by conventions for encoding such things as diacritical marks, non-roman alphabets, and special symbols. As with the descriptive elements of document representations, acceptance of certain conventions for encoding simplifies the task of the system user.

4.2 Query Representations

Because one is restricted to asking questions answerable from available representations, it is not surprising that the representation of queries parallels in many ways the representation of documents. The elements of queries can be descriptive or subject, controlled or uncontrolled vocabulary, with or without syntax. Two major categories of query representations can be identified: those which link terms with boolean operators (AND, OR, NOT) and those which treat the query as a vector of (possibly weighted) terms. In the latter case the query may include specification of a threshold value which must be reached before a document representation is judged to be sufficiently similar to a query representation to warrant its retrieval.

A given query may in fact have more than one representation: an external

representation as input by the system user and an internal representation more suitable for comparison with document representations. For example, a system may allow the user to enter a natural language query which is then translated into an internal representation. This translation process is akin to automatic indexing in the creation of document representations and is a source of variation in internal query representations when different systems adopt different processes. It should also be noted that a query representation is not necessarily a static object. Particularly in interactive systems, the query representation may be modified following a review of preliminary retrieval results.

4.3 Document-Query Relationships

Each time a query is processed, relationships are determined between the query and each document in the database (or a subset of the database). This relationship may be thought of as the system's judgment of the relevance of a document to a particular query or as the degree of similarity between document and query representations. Where queries take the form of terms linked by boolean operators, the relationship has two possible states: relevant or not relevant. Where query and document representations take the form of vectors of weighted terms, the relationship has a range of values and the documents may be ranked with those most similar to the query being retrieved first. Variations in the algorithm by which the system determines the strength of the relationship between the query representation and each document representation can lead to differences in the sets of items retrieved. When systems are evaluated, the correspondence between the system's judgment of relevance and the user's judgment of relevance is assessed.

4.4 Term Relationships

Terms are clearly an important component of document and query representations in many IR systems. A thesaurus stored online is a familiar tool for explicit representation of semantic relationships among controlled vocabulary

terms. Although thesauri are generally compiled by human experts, term relationships can be derived automatically and represented by term clusters. Processes used for deriving term clusters (which assume semantic relatedness among members of the cluster) include both statistical cooccurrence (e.g., Dillon and Caplan) and syntactic analysis (e.g., Hirschman et al.). While in principle term relationships could be identified by processing all of the documents in the database, in practice they are more likely to be derived using a small subset of documents. For example, the new "zoom" feature of the ESA-IRS search system identifies controlled vocabulary terms which frequently cooccur with a particular natural language term in a sample of retrieved document representations. In multidatabase searching another type of term relationship is important, that which links corresponding terms in various controlled vocabularies as exemplified by the vocabulary switching system (VSS) developed at Battelle (Niehoff et al.).

4.5 Document Relationships

Just as the creation of term clusters is a way to explicitly represent relationships among terms, document clusters represent relationships among documents. Documents are judged to be similar if the document representations are sufficiently similar, so the criteria used for establishing document clusters are a function of the underlying document representations. Both shared terms (e.g., Croft) and shared citations (e.g., Schiminovich) have been used as a basis for defining document clusters. Because citations define a relationship between citing and cited documents, representation of these citation links explicitly in the computer would allow one easily to exploit this relationship. Nelson's proposal for hypertext is suggestive of how such links could be followed, moving from citing to cited document and back again when the full text of documents is available for display.

5. NEEDED CATEGORIES

The previous section reviewed those categories of representations which have received the most attention to date. As noted earlier, they include objects (queries, documents), relationships (document-query, term-term, document-document), and processes (for generating document and query representations, for determining the relevance of a document to a query). The dominant categories reflect a preoccupation with systems for document retrieval. As information retrieval comes to be defined more broadly to include question answering as well as document retrieval, additional categories of representations will emerge. This section provides examples of categories which are in need of much more study. In several cases research in AI is suggestive of some of the directions which such study might take.

5.1 Knowledge/information/data

As indicated in section 2 of this paper, knowledge representation has become an active area of investigation within AI. Documents, with which most IR system designers have been concerned, are of course only one possible repository of knowledge to be represented. Two others are human experts and collections of data. In the context of medicine, for example, it is necessary to determine how to represent the clinical expertise embedded in patient records (e.g., Slamecka et al.). Rich (1983) provides a helpful overview of approaches to knowledge representation in AI, distinguishing three categories: knowledge representation using predicate logic, knowledge representation using other logics, and structured representations of knowledge. As IR systems encompass objects in addition to documents, categories of representation will have to be augmented to include alternative forms of knowledge representation. Similarly more variations in query representations are likely to arise when systems are designed to answer questions using processes not limited to term matching. In such systems a useful taxonomy of questions would have categories which could

predict the processes needed to understand and answer the questions falling in that category (Lehnert).

5.2 Complete Documents

Even if one's focus of attention is restricted to documents, it is evident that available representations are rather limited in that they emphasize words and disregard the importance of graphics and pictures as aspects of document content. There is a need to investigate ways of incorporating these in document representations. Although there is growing availability of electronic journals with the full text of articles online, only those graphics which are easily formatted for input and output are included. As input and output devices become less restrictive, opportunities to couple illustrations with the textual portion of document representations should arise. It is possible that much of what has been said concerning document representations in this paper is too conditioned by familiarity with existing IR systems and the documents to which they provide access. As the computer is more widely used for creating documents, document representations may change considerably.

Licklider accounted for the widespread acceptance of IR systems with document representations based on sets of terms by noting that they seem to be effective for document retrieval and even in the retrieval of relevant passages within documents. He suggests that their shortcomings are not likely to manifest themselves strongly until an effort is made to infer consequences from the stored representation of knowledge. Thus as system objectives change from document or passage retrieval to retrieval of information which may be implicit in the text, new forms of document representation will be needed. The IR literature does contain a few suggestions of alternatives, such as Strong's algorithm for generating structural surrogates of English text, but for the most part this aspect of document representation remains to be explored.

5.3 Databases

At present in IR systems providing access to multiple databases, each database is treated as a separate entity and the user must select one or more to which to address a query. This suggests a need to develop representations for databases which could be used in determining the database(s) best suited for a particular query. In commercial systems which make available a "database of databases" to support database selection (e.g., DIALINDEX on DIALOG), the representation of each database is its inverted file, in essence the union of the document representations. In contrast to this highly detailed representation, Marcus is exploring the possibility of representing each database by one or more appropriate class codes from a general classification scheme. The need for database representations will increase as the number of databases online proliferates.

5.4 Users

Van Rijsbergen has observed that "a major stumbling block in designing effective IR systems is the uncertainty about how to incorporate knowledge about users into our system design." (p. 265) At present most systems are neutral: given a query, the same response will result regardless of who has submitted that query. Implicit in this design is the assumption that all users can be represented in the same way, that the interaction and the particular response can be standard for all users. Although this assumption may be reasonable where the user group is relatively homogeneous and well trained, it becomes questionable when a more heterogeneous group of users can be anticipated. Hollnagel and Woods have recently suggested that the user representation should not only be matched to the user's cognitive characteristics, but also be dynamic. This is because a user may change over time, because the nature of the user's task may change dynamically, and because there may be different populations of users. Rich (1979) has proposed a preliminary

taxonomy of user representations: for canonical vs. individual users, constructed by the user himself vs. by the system, containing short term specific information vs. longer term more general information.

The possibilities for user representation have been explored to a limited extent in experimental IR systems. Most work has been done in the area of relevance feedback, where the user's relevance judgments on a sample of retrieved documents are used by the system to modify the set of items retrieved in order to better match the user's interests (e.g., Noreault). This illustrates a form of user representation emphasizing individual users and short term specific information. As Rich's taxonomy of user representations suggests, there are many other approaches which could be explored in the context of IR.

Another interpretation of the concept of user representation would be to represent the experience of users of the system in a form that could be helpful to subsequent users. Representations in this category include sets of terms used together in queries and user comments on items retrieved. Each of these has been investigated in the context of experimental systems. SUPARS maintained a file of query representations, sets of terms which were related in the sense that they had been searched together by a previous system user (Atherton et al.). This is one technique for establishing representations of term relationships. Intrex, as part of its detailed document representations, included a place for user comments which were then accessible to subsequent users of those document representations (Overhage and Reintjes). Both these forms of user representations are dynamic in that they are augmented over time.

5.5 Scope of the System

As long as system designers can assume that users have received training in system use and are aware of the set of queries to which the system can respond, there is no need to be concerned with ways of representing the scope of the system. But if a system is to be used by naive users with no prior

formal training, there is a need to develop mechanisms for representing and communicating the scope and limits of a system. This issue has been dealt with in some depth by Buchanan (1979) in the context of intelligent assistant systems. In order to use a system effectively, a user needs to understand the system's scope and limits. Buchanan defines scope as the broad class of problems which the system is designed to solve and the limits as the idiosyncrasies that must be remembered to obtain reliable solutions. He suggests several ways in which systems could represent their own problem-solving abilities and explain them to users. Just as this area of representation has assumed increasing importance in AI with the development of intelligent assistant programs, it is likely to be of growing importance in IR as systems diversify in scope and user groups become more heterogeneous.

6. EVALUATION

As suggested in section 3 of this paper, a taxonomy should provide a framework within which alternatives can be evaluated. The categories of representation described in sections 4 and 5 reveal that there are several alternatives in need of comparative evaluation. It remains to propose the evaluation measures which could be used. Although the scope of our study does not include evaluation of the various representations identified, we are taking note of the evaluation measures which have been used in IR experiments. Most IR evaluation studies to date have emphasized system performance measured in terms of recall and precision. A study conducted at Syracuse University indicates that reliance on only these measures can mask differences between representations (Katzner et al.). Results showed that for any two representations considered, values of recall and precision differed only slightly while overlap scores were also low. Thus the use of different document representations led to the retrieval of very different sets of documents. The results of this study suggest that equivalence of representations should be decided on the basis of

similarity in performance and similarity in documents retrieved.

Recall and precision are aggregate performance measures useful for evaluation of the system as a whole. For evaluation of the various representations which make up an IR system, there is a need for additional measures. The work on software metrics in computer science is one source of ideas. As Denicoff and Grafton observe: "Metrics for the established science and engineering disciplines are based in the physical sciences, while those for computer science, like software itself, will be derived from human ingenuity." (p. 13) Gilb's monograph on software metrics suggests some measures which could be applied in IR, such as tolerance (the ability of a system to withstand a degree of variation in input without malfunction or rejection) and adaptability (the ease with which new functions can be added to a system). Gilb also notes the importance of what he terms resource metrics (e.g., cost, time, computer and human resources required to accomplish the task for which the system is designed). In the evaluation of AI systems the concept of representational adequacy has been used (Rich, 1983). This is the ability of the chosen form of representation to represent all of the kinds of knowledge that are needed in the domain of interest. For example, representational adequacy in the context of query representations might be termed "habitability" (Watt).

Although there have been a number of evaluations of computer-based IR systems, there is still much to learn. Studies to date have been limited in at least two respects: (1) there has been much attention given to recall and precision to the exclusion of other evaluation measures and (2) only a few studies have done comparative evaluations of alternative representations within a particular category (e.g., documents, queries). As a result there is little guidance for the designer who wishes to choose a particular form of representation best suited to a given situation. The use of multiple measures will likely reveal tradeoffs which the designer should take into account in choosing

one form of representation over another. We may find that different approaches are not simple alternatives, but are best suited to accomplish somewhat different ends.

7. CONCLUSION

This paper has outlined a taxonomy of representations in IR system design derived from an examination of the IR literature. As such it is an initial attempt to codify what has been done, but at the same time it reveals much that remains to be investigated. Additional needed categories of representations and types of evaluation studies have already been enumerated. Two other points worthy of mention deal with mental representations and description of taxonomic category members. With respect to mental representations, we note that their exclusion from our study should not be interpreted to mean that we view them as unimportant. In fact one approach to developing more effective computer-based representations may be to gain a better understanding of the mental representations of system users. With respect to description of taxonomic category members, we note that the natural language descriptions given in this paper should be supplemented by more formal descriptions wherever available. In particular the efforts to develop mathematical models of IR systems are applicable here.

The recent computer science literature offers neologisms such as knowledge engineering (Tou), information engineering (Wegner), and cognitive systems engineering (Hollnagel and Woods). To determine the relationships between these fields and information retrieval, one must study not only how these fields are defined by their proponents, but also how information retrieval has been defined by IR system designers. The taxonomy presented in this paper provides one view of the latter and thus serves as a tool to summarize what has been accomplished as well as to suggest what more could be done.

8. ACKNOWLEDGEMENT

This work was partially supported by the National Science Foundation, Division of Information Science and Technology, under grant NSF-IST-8208576.

9. REFERENCES

- Anderson, G. A.; Jensen, C. D. 1975. Computer interconnection structures: taxonomy, characteristics, and examples. *Computing Surveys*. 1975 December; 7(4):197-213.
- Atherton, P.; Cook, K.H.; Katzer, J. 1972. Free Text Retrieval Evaluation. Syracuse, NY: Syracuse University, School of Library Science; 1972. (RADC TR-72-159, AD 748 218)
- Buchanan, B. G. 1979. Issues of representation in conveying the scope and limitations of intelligent assistant programs. In: Hayes, J.E.; Michie, D.; Mikulich, L.I., eds. *Machine Intelligence 9*. Chichester: Ellis Horwood; 1979. 407-425.
- Buchanan, B. G. 1982. New research on expert systems. In: Hayes, J. E.; Michie, D.; Pao, Y. H., eds. *Machine Intelligence 10*. Chichester: Ellis Horwood; 1982. 269-299.
- Chomsky, N. 1980. *Rules and Representations*. New York, NY: Columbia University Press; 1980.
- Conrath, D. W. 1980. Evaluating the need for burotique: some taxonomic and methodology issues. In: Naffah, N., ed. *Integrated Office Systems--Burotics; Proceedings of the IFIP TC-6 International Workshop on Integrated Office Systems--Burotics*. New York, NY: North Holland; 1980. 199-208.
- Cremmins, E. T. 1982. New terms for old: Information and knowledge representation. *Bulletin of the American Society for Information Science*. 1982 October; 9(1):44.
- Croft, W. B. 1977. Clustering large files of documents using the single-link method. *Journal of the American Society for Information Science*. 1977 November; 28(6):341-344.
- Denicoff, M.; Grafton, R. 1981. Software metrics: a research initiative. In: Perlis, A.; Sayward, F.; Shaw, M., eds. *Software Metrics: An Analysis and Evaluation*. Cambridge, MA: MIT Press; 1981. 13-18.
- Dillon, M.; Caplan, P. 1980. A technique for evaluating automatic term clustering. *Journal of the American Society for Information Science*. 1980 March; 31(2):89-96.
- Fodor, J. A. 1981. *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press; 1981.
- Gilb, T. 1977. *Software Metrics*. Cambridge, MA: Winthrop Publishers; 1977.
- Hirschman, L.; Grishman, R.; Sager, N. 1975. Grammatically-based automatic word class formation. *Information Processing and Management*. 1975; 11(1/2): 39-57.

- Hollnagel, E.; Woods, D. D. 1983. Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Studies*. 1983; 18(6):583-600.
- Jeffrey, C. 1973. *Biological Nomenclature*. London: Edward Arnold; 1973.
- Katzer, J.; Tessier, J.; Frakes, W.; Das-Gupta, P. 1983. A study of the overlap among document representations. *ACM SIGIR Forum*. 1983 Summer; 17(4):106-114.
- Kent, W. 1978. *Data and Reality: Basic Assumptions in Data Processing Reconsidered*. New York, NY: North Holland; 1978.
- Kerschberg, L.; Klug, A.; Tsichritzis, D. 1977. A taxonomy of data models. In: Lockemann, P. C.; Neuhold, E. J., eds. *Systems for Large Data Bases: Proceedings of the 2nd International Conference on Very Large Data Bases*. New York, NY: North-Holland; 1977. 43-64.
- Lancaster, F. W. 1977. Perspective paper: information science. In: Walker, D. E.; Karlgren, H.; and Kay, M., eds. *Natural Language in Information Science: Perspectives and Directions for Research*. Stockholm: Skriptor; 1977. 19-43.
- Lehnert, W. G. 1978. *The Process of Question Answering: A Computer Simulation of Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1978.
- Licklider, J. C. R. 1965. *Libraries of the Future*. Cambridge, MA: MIT Press; 1965.
- Marcus, R. S. 1980. Search aids in a retrieval network. *ASIS Annual Meeting Proceedings*. 1980; 17:394-396.
- McKenzie, K. I.; Spiers, J. D. 1979. Towards a taxonomy for information systems. *Information Privacy*. 1979 January; 1(3):116-119.
- Nelson, T. H. 1967. Getting it out of our system. In: Schechter, G., ed. *Information Retrieval: A Critical View*. Washington, DC: Thompson; 1967. 191-210.
- Niehoff, R. T.; Kwasny, S.; Wessells, S. M. 1979. Overcoming the data base vocabulary barrier: a solution (vocabulary switching). *Online*. 1979 October; 3(4):43-54.
- Noreault, T. 1979. *User Directed Relevance Feedback*. Syracuse, NY: Syracuse University, School of Information Studies; 1979. (Ph.D. dissertation)
- Overhage, C. F. J.; Reintjes, J. F. 1974. Project Intrex: a general review. *Information Storage and Retrieval*. 1974 May-June; 10(5/6):157-188.
- Raphael, B. 1976. *The Thinking Computer: Mind Inside Matter*. San Francisco, CA: W. H. Freeman; 1976.
- Rich, E. 1979. User modeling via stereotypes. *Cognitive Science*. 1979 October-December; 3(4):329-354.

- Rich, E. 1983. Artificial Intelligence. New York, NY: McGraw-Hill; 1983.
- Schiminovich, S. 1971. Automatic classification and retrieval of documents by means of a bibliographic pattern discovery algorithm. Information Storage and Retrieval. 1971 May; 6(6):417-435.
- Simon, H. A. 1981. The Sciences of the Artificial. 2d ed. Cambridge, MA: MIT Press; 1981.
- Simon, H. A.; Siklóssy, L. 1972. Representation and Meaning: Experiments with Information Processing Systems. Englewood Cliffs, NJ: Prentice-Hall; 1972.
- Siamecka, V.; Camp, H. N.; Badre, A. N.; Hall, W. D. 1977. MARIS: a knowledge system for internal medicine. Information Processing and Management. 1977; 13(5):273-276.
- Sparck Jones, K. 1981. Introduction. In: Sparck Jones, K., ed. Information Retrieval Experiment. London: Butterworths; 1981. 1-5.
- Strong, S. H. 1974. An algorithm for generating structural surrogates of English text. Journal of the American Society for Information Science. 1974 January; 25(1):10-24.
- Tou, J. T. 1980. Knowledge engineering. International Journal of Computer and Information Sciences. 1980 August; 9(4):275-285.
- van Rijsbergen, C. J. 1983. Information retrieval: new directions: old solutions. ACM SIGIR Forum. 1983 Summer; 17(4):264-265.
- Watt, W. C. 1968. Habitability. American Documentation. 1968 July; 19(3): 338-351.
- Wegner, P. 1982. Paradigms of information engineering. Knowledge: Creation, Diffusion, Utilization. 1982 December; 4(2):227-243.

How to know what is known.

Designing Crutches for Communication

1. Why to know?

Some people want to know something what they do not know. Why? There are several reasons. One is pure curiosity - which may be rooted in the subconscious suspicion that the unknown item might be somehow useful for some future situation they could encounter. Or: the missing item might help to balance some discrepancy in one's picture of the world (and thereby help to contribute to one's peace of mind), or to brighten some dark spot in one's picture of the world (and, therefore, again help to pacify one's mind). Or: They just LIKE to get new knowledge, without intent to USE it for immediate practical purposes (PURE, entertaining curiosity!).

Other people want to know because they have a problem: they do not know what to do next. And - whatever they do next - should be based on some knowledge of the state of the world, of the world as is ought to be, of the workings of the world and of the ways of manipulating the world (or, at least, some of its parts).

Those processes which lead to this change of somebody's knowledge are called INFORMATION. Information is an event resulting in the change of somebody's knowledge. It is not something "stored" in paper documents, or in RAM files, or the like - although ink blots on paper or holes in a punch card or phosphorescent patterns on a vacuum tube may - occasionally - trigger information.

Where and how to obtain this information?

There are several ways of trying to acquire desired knowledge:

- further thinking and contemplation might lead to the desired certitude;
- gathering first-hand evidence by inspection (observation, measuring, experimenting);
- asking somebody else who is suspected to know;
- asking somebody else whether he/she knows somebody who might know;
- trying to find an answer by browsing/ scanning through/ looking up "documented knowledge" - in libraries of books, or in computer based collections of 'document-like code';
- he/she might take a - more or less 'educated' - guess.

Which of these channels is most appropriate in a particular situation is a most difficult question and a very important information problem of its own right.

Not infrequently, information happens to someone who has not asked for it.

(..... surprise, irritating findings, confusion, ...)

INFORMATION SCIENCE is out to help the information seekers. In order to help/cure others, one needs to know (and this is another knowledge/info. problem) what the "HELPEES" need to know. Do the HELPEES know what they need to know. If not, who else? The HELPER? Perhaps not - if he/she is honest. If not, how can the helping information scientist and information system designer find out?

Designing an information system for somebody else (usually even: a large number of unknown potential information seekers) requires information about the problems of those future (anonymous) "USERS". This task implies not more and not less than "breaking one's head on behalf of future unknown users with unknown problems".

A formidable task.

If information science/engineering does not recognize this task, it will have failed its self-proclaimed mission.

This task presents a number of considerable epistemological and logical problems - in comparison with which hardware and software considerations are almost trivial. The bottleneck is not so much the size and speed of computers or the efficiency of programming but the understanding of KNOWLEDGE AND REASONING. It is not a problem of software engineering but of information engineering - "infware" before software.

The main questions to be dealt with are:

- what is knowledge? which kinds of knowledge should be distinguished? what is their logical status in the processes of reasoning associated with different kinds of problems?
- how is knowledge conveyed, "externalized" in order to preserve it through time or to transfer it to another person? COMMUNICATION is nothing but one mind's attempt to affect some other person's state of knowledge, i.e. to 'trigger information'.
- how is knowledge encoded appropriately in order to guarantee proper decoding?
- what are the weaknesses of natural intelligence? where - if anywhere - can it be supported and how?
- how to identify a potential source of knowledge? How to describe the "content" of a brain, a library, a book? how to know what is known? If anywhere, where is this knowledge to be found?
- which operations of reasoning can be modeled in a way that they can be "delegated" to an external algorithmic mechanism?

The theoretical core of the information sciences consists largely of the knowledge responding to these questions. Whatever can be PRACTICALLY accomplished, is limited by the state of their treatment. Fortunately: modest as they may be, there are some tools and methods to analyze and to describe problem solving behavior.

(.....)

The epistemological position of the information sciences is further complicated by another fact: information science itself is an information process; therefore it becomes a part of those processes it is out to study, to support, to design. The classical separation between the researcher and his/her object is not tenable. The consequence is the loss of neutrality and immunity which the "normal" scientist or technologist appreciates so much. The information scientist becomes just another agent in the field of numerous, frequently controversial interests. He/she is a party in this field who wants to know "in order to .." - intending to bring about changes that affect others who, of course, react to this endeavour: Knowledge is power.

This situation has methodological implications. It requires a very symmetrical relationship between the HELPER and the HELPEE. They become allies - possibly and usually AGAINST others. The ethical implications are obvious.

The information scientists' toolbox consists largely of communication aids - frequently labeled as "systems research", "user analysis", consequence assessment aids, organizational tools, etc. - exactly the same kind of merchandise they intend to sell to their customers.

2. NI vs. AI

A pair of glasses is meant to enhance somebody's eyesight. An automobil enhances somebody's mobility. The glasses do not see instead of you, or on your behalf. Neither does the automobile relieve you from travelling. They are "prosthetic" devices which support, reinforce, enhance some capacity or activity.

In the same way, information systems are prosthetic aids - devices to support activities of the 'unarmed mind'. With other words, they are nothing but crutches for the intelligent mind. They are no SUBSTITUTES for NATURAL INTELLIGENCE (NI) but enhancers.

Each of those activities under the heading of ARTIFICIAL INTELLIGENCE are motivated by a different ambition. They aim at MIMICKING intelligent behavior; their ideal is to design a machine that might even surpass the capacities of NI (like a chess player that beats the world champion). Devices of this kind are SUBSTITUTES and not prosthetic. Occasionally, one can hear or read statements like "Some day the computer will even be able to find its own problems", or "Once we have understood the working of the brain we will be able to design a better one". With other words: this is the ambition to construct the GOLEM, the synthetic homunculus, indicative of the blasphemous desire "to outcreate creation". Many a project of designing so-called expert systems or 'heuristic problem' solvers indicate this attitude.

Of course: a real homunculus is able to exert judgment. It is able TO JUDGE LIKE (or: on behalf of) A HUMAN - perhaps like the right, objective, neutral person. Unfortunately - or fortunately - this seems to be a real and fundamental barrier of what can EVER be delegated to a computer or any other 'algorithmic system'. What is good, desirable, relevant, and important defies - as a matter of principle - any attempt of finding an objective and definitive algorithm that mimicks those processes which produce human judgments. A little reflection reveals the reason: trying to 'found' one's judgment, to arrive at a judgement by deliberation rests ultimately and necessarily on 'offhand judgments', i.e. on judgments which are made 'off-the-cuff' and - therefore - are not justified or justifiable (by deliberation), and which therefore defy algorithmization (which is a prerequisite of computerization). In order to simulate a human problem solver, a program would have to take into account the totality of the human's experiences, knowledge, intentions which it is supposed to mimic. Additional epistemological difficulty: trying to explore somebody's personal knowledge affects that person's personal knowledge. Therefore, what is to be mimicked is already changed by the attempt of mimicking it.

The number of fundamental obstacles against the GOLEM ideal can be extended - in particular, because the pursuit of this ideal is ridden by vicious circles and infinite regresses. Computers are doomed and stuck to algorithmic procedures - even if these are called 'heuristic'.

Looking at it more closely: who WANTS the GOLEM ideal (for automatic indexing or translation, for finding THE OPTIMAL SOLUTION to a practical problem, for a RELEVANT PROBLEM IDENTIFIER)???

(.....)

More promising, friendly and realistic seems to be a less ambiguous program than that ideal of AI. Instead of substituting for human intelligence, judgments, and emotions, let us try to enhance, to reinforce, to guide NI. It seems to be worthwhile to make it less probable that NI falls victim to its 'natural' weaknesses, such as the tendency not to perceive what is conflicting with one's pet ideas and prejudices, to reject irritating news, to neglect the search for long term consequences if the short term payoff promises some advantages, etc. etc.

Unfortunately, some crutches - and these are usually the good and comfortable ones - build up a dangerous temptation of their user. Since they are fun to use, or because they are very easy to use to a limited set of ends, or because taking trodden paths of thinking is the easiest, laziest, and -therefore- the most 'economic' way to get SOMEWHERE -- the user or operator of the crutch is tempted to let the crutch take its own way, its customary, easiest path. The more adaptive and the higher the "learning ability" of the crutch, the greater the effort to retrain the helpful device and the inclination to do want what the crutch can easily do - instead of struggling with the crutch in order to make it help accomplish what one WANTS to do. In these cases, the crutches have changed into an automaton. Oftentimes, the masters do not even realize this change, or - if realized at least subconsciously - the masters deny vehemently that this development is odd, irritating or consequential. On the

contrary: There are many triumphant claims of successes of AI, i.e. of substituting NI by a synthetic device. And if there are some, why objections, whether the NEW TECHNOLOGY might not lead to an oversimplification of the problems to be dealt with, or to the treatment of inadequate problems (just because they are so easy to be dealt with!): IF THE PROBLEMS DON'T LIVE UP TO THE METHOD - THE WORSE FOR THE PROBLEMS!

The recent history of the information sciences is full of examples for this effect. Information systems abound which are being used as YOUR FRIENDLY INTELLIGENCE SUBSTITUTES.
(..... Examples:)

3. Reinforcing NI through Information Systems.

What are the typical properties of a device which is apt to reinforce and to extend NI?

If we speak about a 'device' or a 'system' we do not mean a piece of hardware. It does not even have to contain much or any hardware at all. In any case, however, they are constructs of rules and procedures which are meant to serve the desired end. They may also incorporate hardware, such as books, telephones, mail-boxes, computers. In addition, there may be human operators who "man" the system (telephone operators, documentalists, mailmen, programmers, supervisors, repair-persons - who are faced with their own information problems).

The purpose of an IS to reinforce some person's information, i.e. the change of his state of knowledge:

- by confirming what that person already knows, making him more certain;
- by adding to his knowledge;
- by weakening his knowledge, making him less certain;
- by deleting some item(s) of his knowledge, increasing ignorance.

Most information systems designed these days are of the first and/or second type. They are CONFIRMATION SYSTEMS. They are very likeable and popular because they reinforce the ego of their users and designers. Some information scientists would even go so far to claim that ALL information systems should be of this kind, that they should reduce uncertainty.

Nevertheless, the usefulness of confirmation systems is rather limited in view of NI-reinforcement. Information systems which - at least occasionally - stimulate surprise, trigger perplexity, shake up knowledge, tell what one does not expect or even does not like to know can be extremely useful NI-reinforcers. Doubt is the mother (father?) of invention. And reduced certainty leads to less careless plans and actions. Seeing what one would not see otherwise helps to take into account consequences of contemplated actions which would not have been taken into account otherwise. Being made aware of a flaw in one's pet idea reduces enthusiasm and may motivate the search for other ideas. Being informed about ignorance is a prerequisite for

stimulating the search for new knowledge. Only exposure to the opposite view leads to fruitful reconsideration. Etc., etc.

In this context, it does not matter what the SOURCE of information is. It can be the result of immediate communication between persons. It can be the case of mediated and delayed communication through a document. And, most important, it can be communication between a person and him/herself: confrontation with my ideas of yesterday, becoming aware of inconsistencies, contradictions and fallacies, being reminded of what I would have forgotten otherwise, learning what one does not know (but should): the information system as a (somewhat alienated) 'mirror' of one's understanding.

(.....)

In many cases, it is not at all clear and unambiguous who is the USER and who the provider of the systems. Oftentimes (and these are the interesting cases) there is no clearcut demarcation between donors and recipients. Whoever wants to know 'through' the system is not unlikely to become a potential future provider.

Obviously, the NI-reinforcers need something to reinforce. The most powerful amplifier will not generate a measurable current from a zero input. And NI-reinforcers cannot turn stupidity into wisdom. Ideally, a lasting enhancement of NI results from intelligent use of intelligent NI-reinforcers.

4. Some Design Principles.

These considerations show that the "transfer of knowledge" is not a simple data-transportation procedure. Therefore, an information system which is meant to serve as an NI-reinforcer cannot be sufficiently described and designed as a "classical" data processing system (such as a police record file, or a bibliographical documentation system, or a computerized simulation model). It is rather a communication device which is subjected to a number of design principles.

P 1. Whenever possible, the system should establish direct communication between persons. In this role, the system becomes a referral device. It should grow networks of communication links between users which share similar problems.

P 2. Instead of trying to classify users and sources into clearcut conceptual categories, procedures of the kind "who knows of somebody who might know somebody who ... " should be utilized.

P 3. It should not be attempted to "store" all knowledge ahead of time that might be useful. Whatever the system contains should grow with its use.

P 4. Whatever is "stored" and processed by non-human components of such a system are DATA, not knowledge or information.

P 5. Whenever communication is mediated by the system then a code is needed (a 'representation') which allows to formulate data which 'trigger' the appropriate

knowledge in a recipient, with appropriate clarity and precision.

(.....)

P 6. Whatever data are "contained" in the system should not be forced into a rigid hierarchical classification. Instead they should be linked by networks of relationships which correspond to the various types of similarity relations which guide associative processes. Ideally, the system grows "externalized extensions of the user's associative repertory".

P 7. The system should keep a log of its use. In particular, it should keep a record of the problems in connection with which it has been used. The search for previous, similar problems and answers may lead to hints for treating a present problem.

P 8. The system cannot be better than its designer's knowledge of the structure of the knowledge to be dealt with and its dynamics.

(.....)

Her lies the central task of information science: to develop methods for exploring its users' knowledge and their modes of reasoning, i.e. the systems analysis of the logic of problem solving and information. These are prerequisites for a "theory of mental crutches" and the design of better mental crutches, say, NI-reinforcing information systems.

FROM THE STEREOTYPE APPROACH TO KNOWLEDGE REPRESENTATION
TO A POLYMORPHOUS CONCEPT OF KNOWLEDGE ORGANIZATION

Monika Kallfass,
Thomas Seeger,
Projekt INSTRAT, Freie Universität Berlin

1. The stereotype approach as a working concept
Different aspects of the stereotype concept

In our recent work on knowledge representation and organization we used the stereotype approach first and foremost as a label for or path through different problem levels of the knowledge representation field. In this way "the pictures in our head", as Lippmann described the meaning of stereotypes so charmingly, were rather vague and only to some extent in agreement with the concept of stereotype as treated in Social and Cognitive Psychology, or in Psychiatry, which was one of the first disciplines to take over the "stereotype" concept for investigations on patterns of behaviour, and finally, in Artificial Intelligence.

For each of these disciplines the notion of stereotype includes the general aspect of structuring the world in our mind or, to put it in a more philosophical way, the problem of representing the objective world. This cognitive structuring serves as a main human function: to simplify our perception by generalizing, ordering pieces of information in such a way that one recognizes not details only but a whole. Moreover, these generalizations are potentially social, i.e. can be shared with others.

At least in one of the traditional areas of stereotype research, in Social Psychology the stereotype concept implies more than this cognitive dimension. It is thought of as containing an important portion of emotional painting. In fact we can imagine these emotional parts to be the colours in the picture.

Still within in the field of Social Psychology, there is a strong tendency to equate stereotypes with socially negative phenomena like prejudices, for example, defining the concept rather pessimistically as a consequence. The "grain of truth" in this judgement, to use another typical metaphor for stereotypes is probably the fact that stereotype structures can be very res ist-

ant against knowledge, may in fact lead to the rejection and denial of information.

Taking this aspect into account might cast some shadow on the application of the concept for knowledge representation.

When we did not strictly follow these interesting hints from the stereotype concept, was this due to its quality of generalization (a generalization which is useful and applicable for information providing mechanisms in general), keeping us from continuing our work on one aspect or other mentioned above?

On a more serious note, the following considerations about the stereotype direction in Artificial Intelligence and in Cognitive Psychology or Science were the reasons for us not to restrict our approach too early to some specific stereotype concept.

- In Artificial Intelligence which in Information Science undoubtedly is the main area to be considered for stereotype-like knowledge representation, concepts like the frame theory and thus also semantic networking are particularly useful methods for implementing high-level knowledge structures in micro-worlds.

However, neither the application of these methods to large-scale and highly incoherent knowledge bases nor their linking to other AI operating methods applied to knowledge seem to be developed far enough as yet.

We therefore considered a wider theoretical horizon to start with and to guide the application of these methods to be as useful as the exclusive concentration on implementing structures like frames. The outlines of such a broader framework for knowledge organization, which are mainly derived from the work in our project, will be briefly summarized in the following chapter.

- Looking at Cognitive Psychology now, where at least Piaget's "schema" and to some extent the "Gestaltpsychology" started to guide research into the stereotype direction a long time ago, the investigation of stereotype-like cognitive structures is still a challenge to Information Science. Up to now, results from Cognitive Science had too little influence upon the design of information systems.

Therefore, from a somewhat different angle and as a first conclusion; in the long run the stereotype approach, as we can see it, will mean the development of a theoretical concept, which is equivalent to what is called "field theory" in physics, integrates and links together the different concepts like frame theory from Artificial Intelligence, models about knowledge and memory structuring from Cognitive Science, and concept theories so as to further develop theoretical concepts used for documentary languages, and others. We do trust that those who will finally reach that aim, will also find a much better name for it.

2. The broader spectrum of knowledge organization - some principle considerations

Our theoretical approach to knowledge organization and representation is guided primarily by MONSTRAT, i.e. the architecture of an information providing mechanism developed in the INSTRAT Project and presented with specific considerations at this conference by Hennings/Wersig and Belkin/Windel. For the development and implementation of knowledge representation this concept defines two principles:

- The principle of interactiveness

Within an interactive process of human and/or machine problem treatment the knowledge provided by this mechanism needs to be adaptable to the different stages of the process and therefore must be procedural and variabel. These conditions are rather difficult to meet in many traditional information systems.

- The principle of problem adequacy

The knowledge must be structured in such a way that it can be used for different types and structures of problems.

Although this principle has been a main paradigm of Information Science for a long time, we still have to regard it as a task to be completed.

By merely stating such general and almost oldfashioned principles we have not yet proved their usefulness for new approaches to knowledge representation. But in our perspective which by definition of the Project is a long-term and strategic one, for the technological development, the knowledge field, as it

is undergoing fundamental changes, needs a broader context to be embedded in and the final legitimation by societal needs, if we intend to consider it to be the crucial domain for Information Science in the future.

To summarize and to come back to the principles mentioned above, working on more sophisticated knowledge structures will do a lot of good to Information Science and practice. But these developments, however, should be orientated towards taking the problem treatment task for Information Science even more serious than we managed to do so far.

In such a perspective not only the question of what knowledge really is arises again, but a great effort will be required in both research on problem structures and types and on knowledge representation to link the areas effectively.

We concentrated in the concept to be presented here on the problem of knowledge organization which to us is but one piece of the jig-saw puzzle in the future picture of knowledge research, to use this metaphor a last time.

We therefore briefly indicate a few more problems now which are related to but not necessarily covered by the concept to be presented. Since they are as general as the principles mentioned above and lacking some systematic structure they may provoke objection rather than agreement by experts. We did not solve these problems by working on them to some extent besides our main concept. But as they are becoming more and more central problems in our view of the knowledge area, we hope that we shall be able to understand them better. At least or at last we shall use them to evaluate and revise some aspects of the concept presented later.

Basic theoretical problems

- The specification of knowledge types

Existing typologies of knowledge are neither complete nor specific enough to be used for the analysis and construction of knowledge bases. For a better definition of knowledge types a functional approach seems to be rather useful.

- Knowledge, structure or operation?

Before having definitely finished the discussion about the notion of information, how to understand this triad, which means, getting an understanding of what knowledge is, arises as an even more urgent and difficult question, because it implies considerable practical consequences.

Whereas the distinction between knowledge and structure is made in AI-systems for example (i.e. declarative and procedural knowledge), it is still rather unclear how in human cognitive processes one can distinguish between knowledge and structure and vice versa. What is predominant, structure or knowledge? Or, as a question from a more practical viewpoint: how to put what knowledge in which structure and how to operate on it or them?

- Knowledge: As subjective as objective?

As metaphoric as factual?

As pragmatic as syntactic and semantic?

Similar to the problem above these questions are meant to indicate that our understanding of knowledge needs to be limited and broadened to such an extent that it fits into the requirements that knowledge has to meet within an information system. Maybe we need not care about metaphoric knowledge to be presented in such systems, but we do not know yet, because too little thought has been devoted to these questions so far.

- Methodological problems

Methods for the reconstruction of both problem and knowledge structures
As we said before, the reconstruction of knowledge structures seems to be an extremely difficult and haphazard process. If for the necessary empirical evidence we want to find out, for example, to what extent the structures in problem statements can be reflected in the knowledge needed to solve this problem, we must have more elaborate methods to identify such structures.

- After representing, how about presenting knowledge?

Results of psychological research make it clearly show, that different people with their individual cognitive and perceptive styles make better use of knowledge when it is presented in the right way. But what are the right ways? Are the forms of representing knowledge to be different for certain kinds of presentation?

3. The concept of polymorphous knowledge organization

At a first glance none of the considerations made before seem to have directly influenced the following concept, since it refers to a problem that can be described as primarily pragmatic or practical.

However, as regard the tools to deal with or to resolve this problem, not only the idea of stereotype structures but many of the questions indicated before can be found in different ways throughout this concept.

The problem

When we concentrated on developing a concept for knowledge organization, we referred to a problem that centers around the very question of knowledge representation. This is the problem of how one can improve access to and use of the existing information systems by means of a knowledge concept. Probably it was the MONSTRAT concept that made us decide to design or try to design an instrument which could bring traditional information systems closer to the power of those systems which will be implemented over the next years.

To bring them together, which means neither forgetting the old systems nor the new ones, we need to deal with their difference in access to knowledge. Following the line from the simplest case of reference retrieval systems through so-called free text or full text systems, and information offered via interactive videotex, down to the highly complex knowledge-based AI-systems, the development of knowledge structuring methods can be briefly summarized as the following: From the beginning documentary languages such as classification systems and thesauri served as instruments for retrieving reference knowledge. It may be for that reason, that they were concentrated on the linguistic control, especially on the control of concept/term relations. With ongoing technological support systems with free text or full text retrieval reduced this aspect of control for the benefit of greater variety and simplicity. With the latest knowledge-based systems the shift from retrieval-oriented knowledge structuring to knowledge representing structures in the true sense of the term becomes quite obvious.

From the point of view of a user, this variety or polymorphy in access to the systems means a lack of clarity too, so that it becomes almost impossible to know about and use the different systems effectively at the same time.

If a concept like the one presented is supposed to be a kind of aid to achieve transparency in order to make better use of the knowledge in the systems, the aspect of polymorphy must be defined and identified more precisely.

When we use this notion as we did so far, it stands for a problem rather than for a certain quality, because it is related to the discrepancy in power and efficiency of structuring knowledge which we find in the different systems. Moreover, the polymorphy in the questions and problems that take users to an information system is a problem as well.

In another sense, however, polymorphy seems to be the adequate term to characterize the fact that it is the richness and variety of knowledge structures that account for advantages of the newer systems.

With this assumption we decided to conceptualize polymorphy in its positive meaning as an approach to the requirement of transparency and more effective use of knowledge, as we described the problem before.

Because such an approach is supposed to be the direct opposite of the alternative of a stereotype approach, which we can conceive and also like as a sort of superstructure which is general enough to cover various modes of knowledge structuring, the concepts brings another crucial aspect of stereotypes to light that has not been considered explicitly so far, but should be rather important for further application of a stereotype concept. It is the question of how stereotype structures can imply or exclude the dimension of polymorphy.

Conceptualization and tools

Because the concept concentrates on knowledge organization in accordance with its pragmatic foundation, it is orientated towards a tool which can be described as a sort of polymorphous interface between the knowledge represented in the system and the user's problem statement on the other hand.

To reiterate the reasons why we think that such a tool is necessary: it seems quite unrealistic that most of the existing systems can change their ways of representing knowledge enough within the near future. Nevertheless they have to meet certain requirements, no matter how their knowledge is represented. Furthermore, because it is the technological development that caused the difference in systems, we should make use of this very fact to overcome existing deficiencies.

By analyzing the conditions of the system, which have to be considered for such a tool, we found it necessary to distinguish between three basic aspects of knowledge:

- The representation of knowledge

The knowledge itself is usually incorporated within a unit which we call "episode". The knowledge can be available in a system in very different forms like references, texts, pictures etc.

- The knowledge structure

When using that "episodic knowledge" the system needs representations of knowledge structures. It very much depends on such instruments of representing knowledge structures (they can be documentary languages or knowledge representation languages), how the knowledge stored can be used efficiently.

- Operations

Especially in AI systems there is a great interdependence between operations and structures. For analytical reasons and for a larger independence from special implementations we made a distinction between the two.

From this triple subdivision it becomes quite clear that at least for existing documentary languages the aspect of knowledge structures is the one to be improved most. We therefore neglected the aspect of operations for the time being, although we consider the operations that are mostly used in highly developed systems to be too much based on logic, and so to be rather insufficient as well.

As we indicated briefly at the beginning of this chapter, documentary languages with their specific mode of access to knowledge are able to realize relational structures to a minor extent only which have to be considered as basic representations of knowledge structures. Although the implementation of

these structures is solved well enough by networks as basic representations, the specification of these relations however remain as a theoretical problem.

Throughout the conceptualization of the polymorphy approach and from the background of other investigations in our project we arrived at the following distinction between relations which we only want to indicate here as briefly as the relational problem:

- Formal relations
(The basis of formal network representation)
- Constitutive relations
(In a nutshell, relations are constituted in various ways: empirically, semantically, evaluating individually.)
- Relations of content
- Performance relations (Leistungsbeziehungen)
(to meet specific requirements).

What we gained from this temporary distinction so far was the possibility to include types of knowledge in theory which are usually neglected in information systems (e.g. appraising knowledge), even if these considerations are still tentative. However, as a central part of the polymorphy concept, we should now take a closer look at the performance relations.

By definition, the performance relations were described as relations that are to be constituted within an information system as to meet specific knowledge functions or requirements. Reference retrieval, for example is a function that knowledge has been organized for, and in documentary languages various performance relations such as movements up and down in a hierarchy, or associative paths can be formally distinguished.

By this functional foundation the performance relations automatically lead to the users' part, because it is from their problems that the functions are defined.

If we are able now to specify these performances from the users' view, we have gained the basis for the design of a tool as described at the beginning.

This step was even smaller than one would have expected, because it is the very same aspect of performance features that can be used to describe the requirements made of the system.

From cognitive models and investigations we derived a variety of classes of performances that can be assumed to be achieved by human knowledge organization. Because these classes, reflecting various theoretical approaches, are polymorphous, they are neither exclusive nor to be combined in one specific way, and there is no clearcut distinction between them. In this sense, to us they are levels rather than classes. Having to confine ourselves to a mere listing of the eleven levels we differentiated, we nevertheless would like to stress their increasing complexity from levels one to eleven, and to indicate that level six to eleven in particular imply the principle of stereotype structuring to a great extent.

Levels of performance

- | | |
|----------------|-------------------------------|
| 1. Existence | 6. Closed fields |
| 2. Attribution | 7. Open fields |
| 3. Relation | 8. Horizontal stratifications |
| 4. Definition | 9. Vertical stratifications |
| 5. Paths | 10. Disperse networks |
| | 11. Systemic structure. |

For a first empirical evaluation of the concept we started to elaborate the performance classes one to five out of short texts or "episodes", as we call them, from a student's advisory pamphlet. We thereby used primarily substantial key terms in these texts as description units or entities. Although the reconstruction of these performance structures in the text was not very easy, it gave us empirical evidence that the concept of polymorphous performance levels is in principle practicable and transformable into (maschine)implementation.

As we like to indicate in the following conclusion, however, in order to develop the polymorphous tool that we aim at, these performance levels need further specification.

4. Conclusion

As a summary of our approach presented, we consider two aspects of the polymorphy concept to be crucial:

- The concept stands for the goal to enrich knowledge structures in such a way that they allow multiple access to knowledge.
- Concentrating on the aspect of access to knowledge more than on knowledge representation, the conceptualization of performance levels to us seemed to be the possibility for structuring or representing the users' queries in a polymorphous way.

From the evaluative point of view these performance levels indeed seem to be structures rich enough to reflect the cognitive variety of query structuring and by this means to take the user's side and the problem statement aspect much more into consideration than many existing concepts.

However, besides the necessity for their further specification, we experienced these structures to be difficult to get related to knowledge as represented in texts.

We therefore from our preliminary evaluation at first draw the following conclusion: As presented before, the conceptualization of polymorphous knowledge structuring proved to be of highly analytical value, because it leads from a sufficiently broad theoretical base to a useful differentiation of the knowledge (representation and organization) problem.

As a second conclusion we already indicated the necessity to elaborate the structures already developed to such an extend, that the transformation from an analytical to a constructive concept can be reached.

Finally, as a more general outlook we are confident that the approach of polymorphous or multiple knowledge structuring is broad enough to enable solutions to problems in the knowledge field which are to a greater extend orientated at the users and their problems' aspects.

When we now consider from this point of view the remaining question of how a polymorphous approach could be related to the concept of stereotypes, it becomes clear that the latter can be smoothly embedded in the first in such a

way, that stereotype structures have to be considered as not the only but essential ways of knowledge structuring. Moreover, the design of different and various stereotypes will be an essential task. In addition to this, as our view of the stereotype concept slightly changed through our recent work, we see the necessity not only to distinguish between different forms of stereotypes but between different areas which stereotypes are relevant for: There are knowledge stereotypes as well as stereotypes in the users' problem statements and the queries and they need not necessarily to be the same.

References

Belkin, N.J.; Seeger, T.; Wersig, G.: Distributed Expert Problem Treatment as a Model for Information System Analysis and Design. In: Journal of Information Science, 5(1983), S. 153-167.

Belkin, N. J.: MONSTRAT-Simulation V. Statistical Evaluation of the MONSTRAT Simulation. Projekt INSTRAT, Berlin, April 1983 (A 5/83)

Hamilton, David L.: Cognitive processes in stereotyping and intergroups behaviour. Hillsdale 1981.

Hennings, R.-D. u.a.: MONSTRAT-Simulation III. Register und Materialien. Projekt INSTRAT, Berlin, Juni 1983 (A 3/83)

Hennings, R.-D.: MONSTRAT-Simulation IV. Übergreifende Auswertung. Projekt INSTRAT, April 1983 (A 4/83)

Hewitt, Carl: How to use what knowledge. In IJCAI 4, 1975, S. 189-198.

INSTRAT: Information und Handeln. Orientierungsmuster zur Funktion der Informationstätigkeit für individuelle und gesellschaftliche Problembewältigung. Berlin: Freie Universität Berlin 1982.

Lippmann, Walter: Public opinion. New York 1922.

Seeger, T. (ed.): Information Systems as Collectives of Cooperating Functional "Experts". Informationssysteme als Verbund von kooperierenden funktionalen "Experten". Proceedings of a workshop held in Nov./Dez. 1981 in Berlin. FUB-IFB 5/82, Berlin 1982

Seeger, T. u.a.: MONSTRAT-Simulation I. Das operationalisierte MONSTRAT-Konzept. Projekt INSTRAT, Berlin, April 1983 (A 1/83)

Seeger, T.; Hennings, R.-D.: MONSTRAT-Simulation II. Die Anlage der Simulation. Projekt INSTRAT, Berlin, April 1983 (A 2/83)

Schaff, Adam: Stereotypen und das menschliche Handeln. Wien usw. 1980.

The relevance of Laplace's 'rule of succession' to
knowledge representation

B.C.Brookes (London)

The relevance of Laplace's 'rule of succession' to knowledge representation.

B.C. Brookes (London)

Abstract

Laplace's 'rule' yields a probability distribution of categories whose validity has been questioned, partly because no example of it has been observed 'in nature'. But as the rule is exemplified by the documentation laws of Lotka, Zipf and Bradford, it appears to express a 'natural' law of categorization which has implications for knowledge representation.

1. The empirical laws of informatics: a historical summary.

During the 20th century several empirical quantitative laws have been discovered in various social activities relating to information processes. Of these laws, that first published by Zipf in 1935, based on his statistical studies of the vocabularies of natural languages, is the most widely known. It has also attracted most of the theoretical research on these laws; the search for the exact formulation of this law has from time to time generated heated disputation, though I doubt whether there is one exact formulation of this law.

The Zipf law has been expressed in several ways: the one I cite here was derived by Mandelbrot (1) who based it on principles derived from Information Theory. If the words of a sample of text are ranked in decreasing frequency of occurrence, the frequency of the word ranked r th is given by the frequency-rank distribution

$$g(r) = k/(r + w)^b, \quad r = 1, 2, 3, \dots \quad (1.1)$$

where k depends only on the sample size and w and b are variable parameters.

The first law of bibliometrics was discovered by Lotka (2) in 1926. He noticed that the numbers of scientific authors who contribute precisely n papers to a given scientific topic within a given period can be expressed by

the frequency distribution

$$f(n) = k/n^c, \quad n = 1, 2, 3, \dots \quad (1.2)$$

where the parameter $c \approx 2$.

And in 1935, Bradford noted a regularity in the numbers of journals and papers contributing to a scientific bibliography. Bradford's ambiguous formula was elucidated by Leimkuhler (3) in a form which is equivalent to the frequency-rank cumulated distribution

$$G(r) = k \ln(1 + r/w), \quad r = 1, 2, 3, \dots \quad (1.3)$$

where $G(r)$ is the cumulated sum of relevant papers provided by the r most productive journals, k is the sample size parameter and w is also a parameter.

There are various other empirical laws, including variants of the Zipf law, all expressed in a variety of formulations. When expressed as frequency distributions, however, they are all very skew and approximate to

$$f(n) = k/n^2, \quad n = 1, 2, 3, \dots \quad (1.4)$$

the inverse square law.

Though the sum of the series $\sum 1/n^2$ converges (to $\pi^2/6$) as $n \rightarrow \infty$, the first moment of the distribution (1.4) is $\sum 1/n$ which does not converge. The inverse square law and others close to it therefore evade the analytical techniques of modern sampling and distribution theory because these techniques require distributions to have at least the first four moments. The inverse square law has none.

2. Gaussian and Zipfian statistics

The Russian author Haitun (4) has recently considered 107 reported empirical distributions arising in the social sciences. He shows how stable distributions, defined as those whose convolutions remain of the same form, can be divided into two main types which he calls 'Gaussian' and 'Zipfian'. These are broad descriptive terms which I shall shorten to

'G-type' and 'Z-type' respectively. G-type distributions are characterized by the very weak dependence of their moments on sample size whereas Z-type distributions are characterized by very strong dependence of their moments on the sample size.

Haitun concludes that all the distributions from the social sciences that he has examined are of the Z-type and notes that G-type distributions arise only in the physical sciences. He further argues that techniques based on G-type statistics are 'inapplicable' to the analysis of Z-type distributions. Haitun's conclusions are important: they point to a void in theoretical statistics. As I have argued for some years, we need to develop a new body of theoretical statistics for Z-type analysis comparable with that already developed for the analysis of G-type statistics.

3. Laplace's 'rule of succession'

One of our basic mental activities is to categorize and name the things in our environment in which others share our interest. The things themselves may be physical but the process of categorizing and naming them is a peculiarly human activity. It is a social activity too because the use of a natural language implies that its users categorize and name familiar things in much the same way.

The first to consider the statistics of categorization was the French mathematician Laplace who, in 1774, published a result known as the 'rule' or 'law' or 'principle of succession' which has been much discussed in the literature of induction over the past 150 years. Laplace considered the probability discussion of a set of A-things when they are categorized into the sub-categories $A_1, A_2, A_3, \dots, A_r, \dots$. The derivation depends on the application of the highly dubious 'principle of indifference' which states that if the causes of events are unknown then assume that they are distri-

buted uniformly over the range of possibilities. The advantage of adopting this dubious principle is that it makes calculation possible where otherwise nothing could be done. And so it gets results which can at least be tested.

The main result obtained by Laplace is that when n of the A-things have already been assigned to their subcategories then, if m of the n have been assigned to category A_r , the probability that the next A-thing, the $(n+1)$ th, will also be assigned to that category A_r is $(m+1)/(n+2)$(3.1)

From this result it follows that when n of the A-things have already been categorised then the probability that the next A-thing, the $(n+1)$ th will be assigned to one of the categories already represented is $(n+1)/(n+2)$..(3.2)

From this result it is possible to consider the probability distribution of the whole set of the sub-categories: since $P(n+1) = (n+1)/(n+2)$, we also have $P(n) = n/(n+1)$ and therefore

$$p(1) = 1/2, \quad p(2) = 1/3, \quad \dots, \quad p(n) = 1/(n+1) \dots\dots\dots(3.3)$$

is the probability distribution derived from Laplace's rule.

Philosophers of induction have sometimes said that though this law is of some interest it would be ^{of} greater interest if it were exemplified 'in nature'. By 'nature' even the most humane philosophers appear to mean the 'physical world' and when they contrive simple experiments to test it they rely on the apparatus of classical probability theory - dice, urns, roulette wheels - which I regard as physical. But clearly, in Haitun's terms, it is not a G-type but a Z-type distribution.

And in fact it does arise, if not 'in nature' then in bibliography. The formulation (3.3) was proposed for the well-known ORSA bibliography by the statistician M.G. Kendall (5). And in Appendix I it is shown that the Laplace distribution is the frequency form of the frequency-rank Bradford law (1.3).

As expressed in (3.3) the Laplace law is too inflexible for general use in informetric analysis and I modify it to

$$F(n) = \frac{k}{w} - \frac{k}{w+n} \quad \dots\dots\dots(3.5)$$

where w is a parameter to be evaluated from the data: its value is usually less than 1. I regard this form as derived from a distribution which has the exact inverse square as its probability density function so that

$$F(n) = \int_w^{w+n} \frac{k_2}{x^2} dx = \frac{k}{w} - \frac{k}{w+n} \text{ as above.}$$

In Appendix II the Bradford law is derived from the Laplace law and in Appendix III the Laplace law is fitted to data derived from a book index.

The index of a book is an information retrieval system. It enables the user to locate information about any specific item it lists. If the book index also indicates, as it usually does, how many pages are devoted to each topic listed, the total index also ^{is} a crude weighted representation of the knowledge the book offers.

But not all book indexes yield Laplace distributions. In choosing a book for this example, I first verified that the index was categorized at only one level of specificity before I began to count. A comprehensive treatise on Biochemistry which I also looked at, but rejected for this purpose, has an excellent index which is categorized at three levels of specificity as indicated by the depths of indentation of the index entries. The Laplace law relates to a categorization at only one level at a time.

4. The Laplace law as an exploratory instrument

One outcome of Haitun's papers should be the development of better analytical instruments for the exploration of the quantitative aspects of information phenomena. A distribution, important in the literature, was

first derived by Simon (6) as one of the variants of the Zipf distribution which he derived on various assumptions about the ways in which a Zipf distribution might retain its stability of form as it continues to grow. One of these is usually expressed as the Beta function:-

$$f(n) = m B(n, m+1) \quad \dots\dots\dots (4.1)$$

where m is a variable parameter.

The modified Laplace distribution can also be expressed in this form:-

$$f(n) = w B(w+n-1, 2) \quad \dots\dots\dots (4.2)$$

and it should be noted that when $m = w = 1$ exactly, then both Simon's distribution and the modified Laplace reduce to the unmodified Laplace:-

$$f(n) = B(n, 2) \quad \dots\dots\dots (4.3)$$

Simon's distribution was adopted by Derek de Solla Price who called it the Cumulative Advantage Distribution and applied it to large scale analysis of citation data.

The Simon distribution may have its uses in macro-statistics in which large series are extrapolated to infinity, but it is very clumsy to use as an exploratory instrument which may involve the calculation of partial sums or the enumeration of particular terms. The MLD, with its simpler form and 'telescopic' properties is convenient and versatile. However, when m and w are close to the value 1 (and $m = 1/w$) both distributions often fit data sets equally well. The usual χ^2 -test for goodness of fit is not refined enough for the discrimination of Z-type distributions.

An issue that has not yet been faced is that though theorists continue to assume that categorization is uniformly applied throughout the sets of entities that are sorted out, as though it were some mechanical or physical process, there is abundant empirical evidence to show that, in fact, rare and abundant things are often categorised slightly but significantly differently within the same set. Bradford recognised this possibility by introducing

his concept of the nucleus. He noted that the journals contributing most relevant papers to a bibliography were usually the specialist journals of the topic. His graph shows that these journals did not conform with the Bradford law to which the peripheral journal conformed. In such cases, the Bradford law applies in a hybrid way - the nuclear and the peripheral journals each conforming with a Bradford law but with different parameters. The nuclear effect, however, is observable and its effects are calculable only when the data are expressed in the frequency-rank form.

By any information measure, a frequency-rank distribution is richer in information than the frequency distribution to which it can be reduced. In a Z-type frequency distribution, N elements (e.g. the journals of a bibliography) are distributed over approx. $N^{\frac{1}{2}}$ classes (i.e. over the numbers 1, 2, 3, ...). In the equivalent frequency-rank distribution, approx. $N \ln N$ elements (e.g. the papers of the bibliography) are distributed over the N classes (e.g. the N journals). The frequency-rank distribution therefore offers the richer theoretical form for analysis and especially when the problems of interest concern the richer end of the distribution.

One advantage of the modified Laplace frequency distribution in the exploration of Z-type statistics is its easy transform to and from the Bradford frequency-rank equivalent.

5. The Laplace law and knowledge representation

The first approach to the capturing of knowledge about some scientific topic T would probably call for a comprehensive bibliography to be prepared with the aid of print-out from the relevant data-bases. We would thus hope to obtain a 'complete' bibliography, i.e. one conforming with the 'complete' Bradford formulation which yields approx. $N \ln N$ papers from N journals.

If all these $N \ln N$ papers were collected, read and studied it would be

found that they were highly redundant and partly contradictory. But it would be possible to reduce this redundant literature to one authoritative comprehensive "Treatise on T" - one single volume. The words of this volume would of course conform with the Zipf distribution though, unless it were very carefully controlled in its vocabulary, I doubt whether it would be 'complete' in the Bradford sense - words of the everyday vocabulary are not wholly independent.

The treatise would however require an index. And if this were indexed at one level of categorisation which specified the numbers of pages which gave information about each index entry term, then this index would conform with the Laplace distribution or with its equivalent Bradford, again in the sense of 'completeness', i.e. the N' index entries would refer to approx $N' \ln N'$ pages.

The next step would be to make the treatise on T the machine-readable data-base of an 'intelligent knowledge-based system' (IKBS) accessible to those who sought knowledge on T. The text of the treatise might be read in complete or it might be broken down into 'atomic sentences' which might be entered in some more compact coded and symbolic form. In either case, I have little doubt that, if this process were efficiently done, that the total computer text, linguistic or symbolic, and even down to the basic machine language 'text' would conform with a Zipf/Bradford 'complete' text, i.e. that the m elements of the text would total $m \ln m$ approx. occurrences.

What is it about the function m^m , or about its log, $m \ln m$, that relates to 'completeness'?

If a given corpus of knowledge were represented in more than one way, we have no means of deciding which method would be the 'best' in any definable sense. If the same corpus of knowledge were made the data-base of two different knowledge-based systems claiming different kinds of advantage to the

user, we have no means at present of comparing in any objective way the performances of the systems.

If Haitun is right, and I believe he is, conventional G-type statistics is 'inapplicable' to such problems. And without a critical calculus based on Z-type statistics the pursuit of 'intelligent' forms of information technology, despite their dependence on sophisticated physical sciences, will remain an art rather than the science it pretends to be.

Appendix I. From the Bradford law to the modified Laplace law.

This transform from the frequency-rank Bradford to the frequency Laplace distribution can be derived at various levels of rigour. This is the simplest.

According to the Bradford law, the number of relevant papers contributed by the journal of rank r is

$$g(r) = k \ln(r + w) - k \ln(r + w - 1).$$

This function does not in general yield integral values of $g(r)$ for integral values of r , so assume that the journal yields the integral number m papers when the value of $g(r)$ lies between $(m - 1 + a)$ and $(m + a)$ where $a \leq 1$. Then let $g(r_0) = a$, so that

$$k \ln(r_0 + w) - k \ln(r_0 + w - 1) = a.$$

Solving for r_0 , we have

$$\ln \left\{ \frac{(r_0 + w - 1)}{r_0 + w} \right\} = -a/k$$

from which

$$r_0 + w = (1 - s^{-a}) \quad \text{where } s = e^{1/k}.$$

Let $g(r_1) = a + 1$, then similarly we get

$$r_1 + w = (1 - s^{-1-a}).$$

Then $f(1) = r_0 - r_1 = (1 - s^{-a}) - (1 - s^{-1-a})$ and since k is large enough we have $s^{-a} = e^{-a/k} \approx 1 - a/k$ so that

$$f(1) = \frac{k}{a} - \frac{k}{a+1} \quad \text{and so on.}$$

Appendix II. From the modified Laplace to the Bradford law.

The density function of the MLD is

$$df(x) = \frac{k}{x^2} dx \quad \text{over the interval } \underline{w} \text{ to } \underline{w+n}. \text{ (II.1)}$$

Now put $y = 1/x$ so that $dy = -1/x^2 dx$ and

$$df(y) = -k dy, \text{ over the interval } 1/w \text{ to } 1/(w+n).$$

$$= k dy \text{ over the interval } 1/(w+n) \text{ to } 1/w. \text{ (II.2)}$$

The order of the inverted scores is thus reversed and so the scores now conform with the required rank order. (The uniform distribution is also of some interest.) To get the Bradford distribution we have to cumulate the scores successively, so we now invert the scores again but this time they are summed to give

$$\begin{aligned} G(y) = F(y) &= \int \frac{k}{y} dy \quad \text{from } 1/(w+n) \text{ to } 1/w \\ &= k \ln (1 + w/n) \quad \text{i.e. the Bradford law.} \end{aligned}$$

These derivations are crude, but more sophisticated derivations are possible and empirical evidence of their equivalence is provided in, for example, (6).

Appendix III. Book-index data and the empirical laws

The data tabulated below state the numbers of pages to which the index entries of a book refer (7).

n	0	1	2	3	4	5	6	7	8	9
f		1109	297	136	65	40	37	32	18	16
1	8	15	14	12	8	5	3	3	-	7
2	5	3	2	2	2	1	2	4	5	1
3	1	1	-	1	2	3	-	1	3	2
4	-	-	1	-	-	-	1	-	-	1
5	-	-	2	-	-	-	-	1	-	1

and other values are: 71,72,80,81, 113,121. Total: 1879

In G-type statistics the fitting of curves is usually done by the use of

moments but in Z-type statistics the method of moments is useless. So, in fitting the MLD to the data above, the simplest way is to take two convenient datum points and calculate the required parameters. Here, taking $f(1) = 1109$ and $F(10) = 1758$ we find $w = 0.6955$ and $k = 1307.6$ and then calculate successive values of $k/(w+r)$ with $r=1,2,3,\dots$. The values of the $f(r)$ are given by the differences. In this way the first 10 values of $f(r)$ are calculated to be 1109 (datum), 286, 131, 75, 49, 34, 25, 20, 17 and 12. Then follow 59 as against 72 found for $f(11)$ to $f(20)$ and 52 as against 49 found for $f(21)$ to $f(121)$.

The calculated figures give χ^2 values well within the 10% limit: for the first 10 values it is 8.214 with the 10% limit for 8 d.f at 13.4; for the 12 values calculated it is 10.745 with the 10% limit for 10 d.f. at 16.0.

The data for the Bradford analysis are obtained by finding the first moment of the frequency distribution given above and then cumulating the sums from the tail towards the origin. The criteria for goodness of fit of the calculated Bradford values to the data need to be established; they are very demanding. In the above frequency distribution the frequency classes ranged from 1 to 121, though in fact they were reduced to 12 because the frequencies were mostly very small and the χ^2 test demands some grouping. In this case, the Bradford data present 1879 degrees of freedom and therefore a much severer test of fit between data and theory.

This need for more discriminating tests for goodness-of-fit has been noted by others also. The fact that the index data conform with the MLD as shown above does not mean that the data do not conform with other possible formulations. The χ^2 -test is much too crude for these purposes and this is one strong reason for my strong advocacy of frequency-rank statistics for Z-type statistics. (7).

References

1. Mandelbrot, B. An information theory of the statistical structure of language. In: "Communication theory (W.Jackson, ed.) London, 1953.
2. Lotka, A.J. The frequency distribution of scientific productivity. J. of Washington Academy of Sciences 16. 317-323. 1926
3. Leimkuhler, F.F. The Bradford distribution. J. of Documentation. 23. 197-207. 1967.
4. Hailun, S.D. Stationery scientometric distributions. Pt I. Scientometrics 4. 5 - 25. 1982. Pt.II. Scientometrics 4. 89-104. 1982. Pt. III. Scientometrics 4. 181-194
5. Kendall, M.G. The bibliography of operational research. Operational Research Quarterly 11. 31-36. 1960.
6. Brookes, B.C. A critical commentary on Leimkuhler's 'exact' formulation of the Bradford law. J.of Documentation. 37. 77-88. 1981.
7. Brookes, B.C. Quantitative analysis in the humanities; the advantage of ranking techniques. In: Studies on Zipf's law (Eds. H Guiter and M.V. Arapov) Quantitative Linguistics. 16. Brockmeyer, Bochum. 1982.

Rainer Kuhlen

Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting

Abstract

The analysis of a set of intellectual abstracting rules taken from actual practice in working systems in West-Germany leads to the conclusion that, so far, there is no way of real machine simulation of intellectual text abstracting capacity, due to the significant differences in machine and human analysis and condensation techniques. But, on the other hand, there is no evidence for the traditional belief that surface oriented text analysis procedures, mainly statistically oriented, can produce high quality abstracting. There is a need for knowledge-based text analysis and processing (inferencing and transformation). As a result of the survey of abstracting rules requirements for an automatic abstracting system are formulated.

1. The scepticism of documentalists towards automatic abstracting
2. Intellectual and automated abstracting
 - 2.1 Survey for abstracting rules
 - 2.2 Practice of abstracting
 - 2.3 Why analyze rules for intellectual abstracting?
3. Results of the survey
 - 3.1 General purpose of abstracting
 - 3.2 General purpose of rules
 - 3.3 Preference for informative abstracts
 - 3.4 Content structure of abstracts
 - 3.5 Formal aspects
4. Results
 - 4.1 Generalizing about intellectual abstracting
 - 4.2 Requirements for an automatic abstracting system

Examples of abstracting rules from the survey

References

1. The scepticism of documentalists towards automatic abstracting

There may have been documentalists in the past who believed in the possibility of a type of automatic abstracting which would produce content analysis of given texts in any desirable form of condensation (cf. Harbeck/Lutterbeck; 1968), but in general the concept of language-dependent intelligence has been reserved for human beings, at least among documentalists. And abstracting belongs to that category of intelligent tasks. The decision as to what should be part of an abstract, to quote a few remarks from actual abstracting rules obtained from a survey (s. below 2.1), "can only be made by persons fully aware of ... research problems, and, if possible, should be made by scientists and engineers" (IRRD: 1982,18) or : "Eindeutige Hinweise ... lassen sich nicht geben; die Auswahl muß deshalb der fachlichen Kompetenz des jeweiligen Referenten überlassen bleiben "(ZFID: 1978,7) And in a letter received on the occasion of the survey we found this particularly sceptical formulation: For the analysis of "technisch/wissenschaftlichen Texten gehört Sachverstand, Berufserfahrung und Kenntnis vom Benutzerbedarf... Der mechanische ...Prozeß (the author refers explicitly to indexing, but text condensation in general is meant, R.K.) kann die Texte immer nur in einer Ebene auswerten, die dem Wert wissenschaftlicher Arbeit oder Industrieerfahrung nicht gerecht wird. Hier ist der Einsatz des Rechners nach unserer Auffassung einmal tatsächlich so gefährlich, wie er in polemischen Darstellungen so oft beschrieben wird"

A research project that aims at building an automatic text condensation system (cf. Hahn/Kuhlen/Reimer:1982) therefore has three choices:

(i) to ignore documental practice (and scepticism) and start building an automatic abstracting system, trusting in the power of text analysis and knowledge representation techniques and their applicability to abstracting.

(ii) to develop a model of automatic abstracting by analyzing a reasonable set of intellectual abstracting rules and by generalizing from documental experience

(iii) to do the first but not abstain from the second, that means, to take what is usable from intellectual procedures for the automatic model and let

what is not usable be a new argument for the thesis that the intelligence of man is a very different thing from the intelligence of machines. This can be a good justification for pursuing an automatic approach and not relying fully on an imitation of intellectual practices.

This paper argues for the third choice. There is some evidence for the assumption that the main components of an automatic abstracting system (knowledge representation and acquisition, knowledge management, inference procedure, and transformational processes) (cf. Hahn/Reimer: 1982) cannot make use of actual intellectual abstracting or abstracting rules, as the case may be. The latter may be of some help for output design and in determining standards of performance. But even here the differences between the two will presumably be greater than the similarities. Machine abstracting will offer new possibilities and limitations altogether.

2. Intellectual and automated abstracting

2.1 Survey for abstracting rules

As part of establishing the conceptual groundwork for the automatic text condensation system TOPIC, which is being developed at the University of Constance (cf. Hahn/Reimer in this section), we have carried out a nation-wide survey among 205 documentation and information organizations which - according to the third edition of the "Verzeichnis deutscher Info-Stellen" (edited by GID, Frankfurt) - actually engage in abstracting. 51 organizations sent back material in the form of rules and examples (12 of which were doubles). 19 answered but didn't have rules in a written form. The percentage of absolute answers was 34%, the percentage of usable answers about 20%. In contrast to the early work of Borko/Chatman (1963), who analyzed 130 abstracting rules, we did not intend to carry out a representative quantitative investigation, but to establish some qualitative arguments as a possible contribution to the modelling process for automatic abstracting. Accordingly the results of the survey are qualitatively sufficient and may be even significant, at least for German documentation centers.

2.2 Practice of abstracting

Abstracting is in general regarded as an art based on experience and subject knowledge. The quality of abstracts depends greatly on the homogeneity of the subject field to be covered and on the abstracter's knowledge of the users' needs.

Abstracts as an information service are still popular, in spite of, and probably because of online retrieval. Results of online searching, poor enough as a consequence of simple keyword or even descriptor indexing and of restricted search facilities with Boolean operators, can be dramatically improved by using abstracts as a filter.

This filter or relevance function is the main purpose of abstracts. The user is to be informed in such a way that he knows enough about the original work to decide whether to study it in more detail or ignore it.

There is some evidence for the assumption that online retrieval will not mean the death of abstracting (although it may mean the death of abstracts in printed form), but will actually raise its standards. That may be the reason why many of the abstracting rules received in the survey are of recent date (cf. appendix).

2.3 Why analyze rules for intellectual abstracting?

With a few exceptions (cf. Gerdel et al:1978) many of the arguments and even the formulations in the rules received in the survey repeat themselves. The German standard DIN 1426 (1973) and the international one ISO 214 (1976) have come into effect. Abstracting is documentationally well controlled. Beyond that there is no real gap in research on traditional intellectual abstracting (cf. monographs by Borko/Bernier (1975) and Koblitz (1972; 1982). Wellisch (1980) presents a bibliography that covers the period from 1876 to 1976 on an international basis. Baker et al (1980) demonstrate the historical development of abstracting based on the example of "Chemical Abstracts", and Skolnik (1979) goes much further back in history to Sumerian times.

But, as already mentioned, there is new activity in the formulation of rules for intellectual abstracting and this development coincides with newer trends in automatic text analysis (cf. Kintsch/van Dijk: 1978) and artificial intelligence (and, perhaps more importantly, with new trends in electronic publishing, which are making scientific full texts increasingly available in machine readable form, cf. Kuhlen: 1983). This may be reason enough to think again about the possible connections between intellectual and automatic abstracting.

According to recent publications it may be argued that question-answering, which was the paradigm of computational linguistics and artificial intelligence in the 1970s is being replaced by full text analysis and text condensation (cf. Habel/Rollinger/Schmidt/Schneider: 1980; Schank/Kolodner/DeJong: 1981; Marsh/Sager: 1982; Hahn/Kuhlen/Reimer: 1982; Hobbs/Walker/Ansler: 1982; Kuhlen: 1983). Some interesting work is being done on a kind of automatic abstracting which is mainly knowledge based (cf. Taylor/Kruse: 1977; Fum/Guida/Tasso: 1982; Hahn/Kuhlen/Reimer: 1982; Schank/Kolodner/DeJong: 1981). And there is still a continuing research tradition of abstracting which is mainly based on statistical means and on extracting techniques aided by surface language analysis (for a short survey cf. Kuhlen: 1983). These statistical and surface-oriented approaches - without making any comment on their usefulness - are not discussed in this paper because they are not likely to obtain standards of performance which would be comparable with intellectual abstracting, itself wholly dependent on knowledge of the field in which texts are to be abstracted.

Knowledge based work in computational linguistics, artificial intelligence, cognitive psychology, and related fields has often not taken information requirements into account. But if automatic abstracting is indeed a real challenge for the scientific fields mentioned above - as it is claimed to be because of "its theoretic and practical implications" (Fum/Guida/Tasso: 1982, p.83) and because of its appropriateness as a model of intellectual performance - then it makes sense to ask as Taylor/Kruse (1977, p. 10) do: "How would an experienced reader go about the task of writing an abstract?" They believe that the answer to this question is the key to reasonable research on automatic abstracting. We are not as optimistic, neither with respect to the possibility of analyzing the intellectual process sufficiently in depth nor with respect to the possibility of transferring intellectual to automated procedures.

Nevertheless, this paper follows these lines on a somewhat broader basis than Taylor/Kruee. In this paper we do not intend to compare intellectual and automated abstracting (cf. Cremmins/Trachtmann: 1981; Popova: 1980), but to put the the emphasis instead on an analysis of rules for intellectual abstracting.

3. Results of the survey

3.1 General purpose of abstracting

Abstracts are made for the "komprimierten, nicht wertenden Wiedergabe des wesentlichen Inhalts eines Dokuments" (BAU: 1980). They should be "eine knappe Zusammenfassung der wichtigsten Tatbestände des Originals" (FIZ-Chemie:1981); as "Textverdichtungen" (text condensation) (BFA-DOK:1979; BAM:1981,1982) they supply criteria in order for the reader to decide on the relevance of the original document (IZDV: 1973; IRRD: 1978; IFAG: 1963; Luft: 1982; ARD/ZDF: 1982; FIZ Technik: 1978; BAM: 1982; ASFA:1978).

3.2 General purpose of rules

It has been often noted that the writing of abstracts is probably based on rules , but these rules have never been written down (be it for reasons of time or of principle). But even people who don't have rules follow in general the guidelines of INIS:1971, which are shared by DIN 1426, according to which the purpose of abstracting rules is "defining the standards... to improve the standard and consistency of abstracts... consistency in information content must be one of the features to be aimed at in abstracting "(INIS: 1971, p. 15). In addition, DIN 1426 argues for consistency because of the increasing use of abstracts in automated systems.

3.3 Preference for informative abstracts

DIN 1426 (3.2.3.1-3) attempts to differentiate systematically different kinds of abstracts (cf. TOPIC:1982, p.19); the discriminating factor is the type of reference to content:

indicative abstract

informative abstract

indicative-informativ abstract

DIN 1426 passes no judgement on these types and refers to them only very tentatively to distinguish types of documents (this is done a bit more concretely by: Energy:1978; INIS:1971; FIZ Technik:1978). According to our survey there is no explicit statement in favour of the indicative abstract; if at all, then only in combination with the informative type (ASFA:1978; BAU: 1980; IRB:1982; ARD/ZDF:1982). These mixed types "geben an, welche fachrelevanten Sachverhalte im Dokument behandelt werden (Indikativer Anteil) und liefern gleichzeitig direkte Information (Ergebnisse Schlußfolgerungen, Hypothesen) aus dem Dokument"(BAU:1980).

Abstracts - and this is a first important assessment of actual practice - are expected to be "as informative as possible" (DKI:1982) (comparable ZPID:1978; Gerdel et al (IDIS):1978; IDIS:o.J. BI-Sportwiss.:1982; VDG-DOK; FIZ-Chemie:1981; Wetter:1978; DECHEMA:1981). Abstracts must inform directly, not indirectly; the "what" is more important than the "about" (BI-Sportwiss.:1982). They should contain concrete results in the form of quantitative or qualitative data (ZPID:1978) and should be "as informative as is permitted by the type and style of the document, that is, they should present as much as possible of the quantitative and/or qualitative information contained in the document" (ASFA:1978, p.1). Sometimes it is argued that the informative abstract should even replace the original (INIS:1971, p.13).

This clear statement in favour of the informative abstract must be put in perspective by remembering that informative abstracts are considered to be most effective in scientific fields where empirical/experimental work with

very specialized topics is the rule. The clear verbal preference for informative abstracts is also somewhat relativized by its co-occurrence with the argument of relevance (cf. BAU: 1980; DKI:1982; Bi-Sportwiss.:1982; ASFA:1978). If the purpose of the information consists mainly in its reference to the original (or in helping the user to decide whether the original may be relevant or not) then the purported informative character is actually more an indicative one.

Another argument for informative abstracts is the requirement that indexing should be possible without access to the original document (Energy:1978; BAU:1980; IZDV:1973; IRRD:1978).

3.4 Content structure of abstracts

The general position of DIN 1426/ISO 5966 needs only slight modification. An abstract should inform the reader about

- a) the subject of publication and the scope
- b) the purpose of the publication
- c) the basic hypothesis, models, and scientific approaches
- d) the methods and technical procedures
- e) new results and theories
- f) conclusions, discussion

The requirements of DIN 1426 are in general accepted, although they are found to be partially contradictory, for example:

- completeness
- exactness
- objectivity
- shortness
- readability

Value judgements, interpretation, supplements, or other modifications of the original are not at all acceptable. Abstracts are required to be objective, even impersonal (IZDV: 1973).

Only very seldom does one find the recommendation to make use of formal text structures. Only (ZPID: 1978) regards structures such as titles, tables of contents, chapter headings, summaries, or reference lists as explicitly useful tools for orientation. But obviously they are often used unconsciously.

3.5 Formal aspects

Besides the influence of online retrieval the format of the common punch card still often determines the length of abstracts. The details given on the required number of characters and blanks vary considerably. But there are three main groups

- a) 100-200 characters (spaces included)
- b) about 1000 characters
- c) 1500-2000 characters

Sometimes the rules recommend an order for the sequence of sentences, such as (IZDV: 1973): first sentence: main subject of the document; second sentence: topics treated; third sentence: methodology and conclusions.

(ZPID: 1978) proposes the general order of scientific texts as an ordering principle: general problem, method, result, discussion (similar UMPLIS:1981). Particularly the first sentence is considered to be significant, because it expresses the subject of the publication (IFAG: 1963; ASFA: 1978). For this purpose even the wording of the title may be used, although, in general, the repetition of the title in the abstract is frowned upon. In the case of mixed forms the indicative part should precede the informative part (cf. ARD/ZDF:1982). In abstracts for chemical work it is the textual part which should precede the structural formulae one.

In order to relate the short proposition of an abstract to the intension and extension of the

original, the use of <modal> expressions, like "short, general, theoretical, thorough, exhaustive", (IFAG: 1963, IZDV: 1973) is sometimes recommended.

The connection with online retrieval has already been mentioned (cf. ASFA:1978; IZDV:1973; VDG/DOK). This is an argument for (BAM: o.J.) to encourage a large variety (!) of abstract expressions to facilitate free-text retrieval.

In general the rules admit that abstracting is more effective the more the abstracter knows about its user. But in spite of this there are only a few recommendations on how to set up user models (ZPID:1978: "durchschnittlicher Psychologe mit Hochschulabschluß"; BFA-DOK:1979; even more tentative INIS:1971; DIN 1426; IFAG:1963; IRRD:1982).

4. Results

4.1 Generalizing about intellectual abstracting

The most important aspect of intellectual abstracting is completely left out of the rules: the understanding of the text and the process of condensation with respect to the anticipated users' interests. The rules legitimately rely on the actual abstracter's knowledge and there is hardly any attempt to formalize or algorithmize the procedure of analysis, understanding, and condensation.

It is hard to imagine that proposals like the following could be of any help for automatic abstracting: "möglichst alles Wesentliche" (IFAG:1963; ZPID:1978; Luft:1982); "kurz, prägnant, inhaltsreich" (BAM:1981; VDG-DOK); "es soll die wichtigsten inhaltlichen Bestandteile und Ergebnisse einer wissenschaftlichen Arbeit in knapper und klarer Form ... darstellen" (Wetter:1982); der Text soll "aussagefähig sein und konkrete Angaben enthalten" (FIZ-Chemie:1981); "a clear condensation of the essential arguments and findings of the original. It should indicate the topic,

methodology, and particularly the results and conclusions obtained, with emphasis on quantitative aspects" (INIS:1971, p.13).

Even the very few bits of heuristic advice are not really useful (useful with respect to a model for automatic abstracting): "welche Erkenntnisse können aus dem Dokument gewonnen werden"; "was wird gezeigt, dargelegt, bewiesen, beschrieben, gefordert?" (BAU:1980). "Was war das Ziel der Untersuchung, wie wurde untersucht, welche Erkenntnisse wurden gewonnen" (BAM:o.J.); "what has been observed"; what can be inferred?" (INIS:1971)

Without referring to the rules specifically, the following requirements for intellectual abstracting seem plausible:

1) Abstracts should be as informative as possible, that means, provide as much quantitative and qualitative data as possible and should show the new scientific results and conclusions.

2) Abstracts claim documental objectivity, that means, the abstracter should be as invisible as possible.

3) Different levels of information are highly desirable, the gradation going from the main theme to very detailed data.

4) A further processing, mainly with respect to indexing and free text retrieval, is desirable.

5) Abstracts should be structured either according to formal ordering principles or to content oriented check lists. The structure should normally reflect pre-analyzed or anticipated users' interests

6) Abstracts should in general not substitute for the original full text. They should lead to the original text by means of the relevance decision.

4.2 Requirements for an automatic abstracting system

Abstracting, as we have pointed out, is an intellectual art and as such not directly transferable to automatic procedures. Cognitive psychology and artificial intelligence have, so far, not provided us with sufficient knowledge about the processes really going on in abstracters' minds when they understand texts and condense them. Thus the direct imitation of an intellectual procedure such as abstracting seems to be out of reach.

And even the other way for a machine to prove its intelligence - by demonstrating comparable output performance - is not likely to be achieved, due to the complexity of the task of synthesizing a new piece of real text by deriving it from formal knowledge structures.

It seems more reasonable for machines not to try to compete with human beings directly. The real chance of automatic abstracting lies in its flexibility. This means, that the output need not be a text but may vary in form over a whole range of possibilities, from a single concept (comparable to a generic classification term) which may be derived as a super-node of a textual semantic network, to selected parts of the original text (comparable to full text retrieval). The advanced zooming techniques of graphic software, supported by appropriate terminal hardware, will contribute to the realization of cascaded abstracts in their gradation from very broad, generic, to very specific, concrete information.

On an abstract level we can formulate the following recommendations for the construction of an abstracting system

- 1) It should be knowledge based and full text oriented, that means it should be able to transfer the text completely into a knowledge structure and to partition the text into its main and subordinate parts.
- 2) It should be able to compare text information with the general standard in the field to be covered in order to identify really new results.
- 3) It should be flexible with respect to different users' needs, that means, the output format should vary significantly both formally and with regard to the content.

- 4) The abstracting system should be embedded into a retrieval system on a full text basis, that means, the highly desirable gradation from very general information about the text to very specific information from the text itself should be realized in one single system.

The two projects under development at the University of Constance, TOPIC and TOPOGRAPHIC, are intended as steps along the way to fulfilling these requirements. TOPIC is responsible for providing techniques of knowledge representation on a frame basis (cf. Hahn/Reimer:1983) and full text parsing (word experts), while TOPOGRAPHIC (cf. Heese:1983) is intended to further the process of producing cascaded abstracts by means of graphical interactive procedures.

Examples of abstracting rules from the survey

(American standard: 1969)

Proposed American national standard for writing abstracts. ISO/TC 46 (USA-6)
920 E, October 1969.-

(ARD/ZDF: 1982)

Anleitung zur inhaltlichen Erschließung von Fernsehproduktionen, 1982
(verbindliche Arbeitsgrundlage für alle Anstalten der ARD und des ZDF).-

(ASFA: 1978)

Guidelines for the preparation of abstracts for publication in 'Aquatic
Sciences and Fisheries Abstracts' (ASFA). Fishery Information, Data and
Statistics Service. Food and Agriculture Organization of the United Nations,
Rome, May 1978.-

(BAM: 1981)

Richtlinien für die Literatúrauswertung im FIZ5-Fachinformation Werkstoffe.
Bundesanstalt für Materialprüfung, Geschäftsstelle Vor-FIZ5. B.
Referierrichtlinien, Berlin, Oktober 1981 (used also by 'VDEH - Verein
deutscher Eisenhüttenleute' and 'GDMB - Gesellschaft deutscher Metallhütten-
und Bergleute').-

(BAM: 1982)

Richtlinien für das Anfertigen von Referaten. Bundesanstalt für
Materialprüfung (BAM). Dokumentation Zerstörungsfreie Prüfung, Fachgruppe
6.2. Berlin, September 1982, 3pp.-

(BAM: o.J.)

Hinweise für die Anfertigung von Referaten für die Jahresberichte über
Holzschutz. Bundesanstalt für Materialprüfung (BAM), Fachgruppe 5.1
"Biologische Materialprüfung", 2pp.-

(BAU: 1980)

Richtlinien für das Referieren von Fachzeitschriften. Bundesanstalt für
Arbeitsschutz und Unfallforschung (BAU). Dortmund, Mai 1980.-

(BI-Sportwiss: 1982)

Richtlinien zur Erstellung von Referaten für die 'Sportdokumentation'.
Bundesinstitut für Sportwissenschaft, Köln, April 1982, 1p.-

(BPA-DOK: 1979)

Arbeitsanweisung für die formale Erfassung und inhaltliche Erschließung für
das "Zentrale Dokumentationssystem" (BPA-DOK. Presse- und Informationsamt
der Bundesregierung. Referat II 4, Bonn, December 1979.-

(DECHEMA: 1981)

Richtlinien zur Literatúrauswertung. DECHEMA: Fachabteilung
IuI/Informationssysteme und Datenbanken, Informationssystem Chemische Technik
Frankfurt, July 1981, pp. 1-8.-

(DIN:1426)

DIN:1426: Inhaltsangaben in der Dokumentation. Berlin:Beuth Verlag 1973.-

(DKI: 1982)

Regeln zur Inhaltsangabe durch Kurzreferate. Deutsches Kunststoff-Institut,
Dokumentation, October 1982, 1p.-

(Energy: 1978)

Energy information data base. Guide to abstracting and indexing. Department of Energy, Technical Information Center, Version February 1978 (available from NTIS).-

(Film und Bild: 1978)

Fischer, P.; Giner, U.; Kanprad, C.; McWilliams, P.; Tudor, S.: Regelwerk zur formalen Erfassung und inhaltlichen Erschließung von audio-visuellen Medien, AV-Dokumentation, Grünwald: Institut für Film und Bild 1978, 124 pp., + appendix.-

(FIZ Chemie: 1981)

Richtlinien für das Anfertigen von Referaten. GDCh - Abteilung Chemie-Information und -Dokumentation Berlin (since 1982 FIZ Chemie GmbH) Berlin/Leverkusen, 1981, 12 pp.-

(FIZ Technik: 1978)

Richtlinien zur Literatúrauswertung. Fachinformationszentrum Technik e.V., in particular Erläuterung zur Pflichtkategorie 21, Inhaltsangabe, 1978, pp.1-3.-

(Gerdel et al: 1978)

Gerdel, W.; Eisenhardt, O.H.; Nacke, O.; Neumann, R.: Schema der programmierten Abstraktion, redundanzfreien Darstellung und Transformation von Aussagen. Institut für Dokumentation und Information über Sozialmedizin und öffentliches Gesundheitswesen, Bielefeld, Oktober 1978.-

(Guide: 1968)

Guide for the preparation of scientific papers for publication. SC/M/D/5, Paris August 1968, in particular: Guide for the preparation of authors' abstracts for publication, pp. 5-7.-

(Hohenheim: 1978)

Hinweise für die Auswertung von Dokumenten. Universität Hohenheim, Dokumentationsstelle, Stuttgart, August 1978, 3 pp.-

(IDIS: o.J.)

Hinweise für das Abfassen von E-Referaten. Institut für Dokumentation und Information über Sozialmedizin und öffentliches Gesundheitswesen, Bielefeld, 1 p.-

(IFAG: 1963)

Hinweise für die Abfassung von Kurzreferaten. Institut für angewandte Geodäsie, Frankfurt, April 1963, 2pp.-

(INIS: 1971)

Instructions for submitting abstracts (IAEA-INIS-4), Vienna, August 1971, 40 pp.-

(IRB: 1982)

Regeln für die inhaltliche Erschließung von Dokumenten. IRB-Informationszentrum RAUM und BAU der Fraunhofer-Gesellschaft, Stuttgart, May 1982, in particular Kat. 600 Referat, pp. 15-17.-

(IRRD: 1982)

Working Rules. International Road Research Documentation (IERD), Chap. IV.2: Preparation of abstracts. January 1982, pp. 18-20.-

(ISO: 1976)

ISO 214: Documentation - Abstracts for publications and documentation. International Standards Organisation, 1976.-

(ISO: 1982)

ISO 5966. Documentation - Presentation of scientific and technical reports. International Standards Organisation, 1982, 22pp.-

(IZDV: 1973)

Internationale Zusammenarbeit zur Dokumentation über Verkehrswirtschaft (I.Z.D.V.) Arbeitsregeln, 1983, Chap.3 "Abfassung des Kurzreferates", pp. 30-35.-

(Luft: 1982)

Richtlinien zur Literatursauswertung und Erfassung selbständiger und unselbständiger Veröffentlichungen. Dokumentationsring Luftverkehr (BFS, DLH, FDL, FAG) 1979, third edition 1982, 19 pp.-

(UMPLIS: 1981)

Anleitung für die Erfassung und Erschließung von Fachliteratur im Rahmen des Dokumentationsverbundes "Umwelt". Umweltbundesamt, UMLIS, Berlin, June 1981, 9 pp. + appendix.-

(Wetter: 1982)

Hinweise für die Abfassung von Kurzreferaten. Deutscher Wetterdienst. Informationsdienst "Dokumentation Meteorologie" (Mitteilung Sept. 1982) 2pp.-

(ZPID: 1977)

Leitfaden für die inhaltliche Erschließung von Zeitschriftenaufsätzen. Zentralstelle für psychologische Information und Dokumentation an der Universität Trier, compiled by Joachim H. Müller, 1977, supplemented 1978, 22 pp. + appendices.-

(ZPID: 1978)

Richtlinien und Hinweise für das Verfassen von Kurzreferaten über psychologische Information und Dokumentationen. Zentralstelle für psychologische Dissertationen an der Universität Trier, compiled by Günther Reinert & Udo Wolff, 1978, 28 pp.-

References

Baker, D.B.; Horiszny, J.W.; Metanowski, W.V. (1980): History of abstracting at Chemical Abstracts. Journal of Chemical Information and Computer Science 20,4, pp.193-201.-

Borko, H.; Chatmann, S. (1963): Criteria for acceptable abstracts: a survey of abstracters' instructions. American Documentation 14,2, pp.149-160.-

Borko, H.; Bernier, Ch.L. (1975): Abstracting concepts and methods. New York et al: Academic Press, 250pp.-

Cremmins, E.; Trachtmann, M.(1981): Writing good quality human and machine abstracts using the 3 r's of reading, rules, and relationship, in: L.F.Lunin et al (Eds.): The information community. Proceedings of the 4th. ASIS Annual Meeting, Washington, 25.10.-30.10.1981. White Plains: Knowledge Industry Publications, pp.157-159.-

Fum, D.; Guida, G.; Tasso, C.(1982): Forward and backward reasoning in automatic abstracting. Proceedings of COLING82, ed. by J.Horecky. Academia: North-Holland Publishing Company, pp.83-88.-

Habel, C.U.; Rollinger, C.R.; Schmidt, A.; Schneider, H.J.(1980): A logic-oriented approach to automatic text understanding, in: L.Bolc(ed.): Natural language based computer systems. München: Hanser, pp.57-117.-

Hahn, U.; Kuhlen, R.; Reimer, U.(1982): Konzeption und Aufbau des automatischen Textkondensierungssystems TOPIC. TOPIC-1/82, 77pp.-

Hahn, U.; Reimer, U.(1982): Informationslinguistische Konzepte der Volltextverarbeitung in TOPIC. Text-Parsing; Frame-basierte Wissensrepräsentation; Kondensierungstransformationen und grafische Visualisierungstechniken. Universität Konstanz, TOPIC-2/1982, 44pp.-

Harbeck, R.; Lutterbeck, E. (1968): Inhaltsangaben in der Dokumentation. Nachrichten für Dokumentation 19, Heft 2.-

Heese, C.(1983): Die Konzeption einer interaktiven graphischen Mensch-Maschine-Schnittstelle für das System TOPIC: TOPOGRAPHIC (TOPIC Operating with GRAPHical Interactive Components). Universität Konstanz, TOPIC-4/83.-

Hobbs, J.R.; Walker, D.E.; Ansler, R.A.(1982): Natural language access to structured text. Proceedings of COLING82, ed. by J.Horecky. Academia: North-Holland Publishing Company.-

Kintsch, W.; van Dijk, T.(1978): Toward a model of text comprehension. Psychological Review 85, pp.363-394.-

Koblitz, J.(1972): Referieren von Informationsquellen. Lehrbrief. Berlin, 174pp.-

Koblitz, J.(1982): Bearbeitung und Verarbeitung von Fachinformationen. Leipzig:Bibliographisches Institut, 304pp.-

Kuhlen, R.(1983): Volltextanalyse zum Zwecke des Abstracting. Proceedings of Deutscher Dokumentartag, Travemünde, 27.-30. Sept. 1982.-

Marsh, E.; Sager, N.(1982): Analysis and processing of compact text. Proceedings of COLING 82, ed. by J.Horecky. Academia: North-Holland Publishing Company, pp.201-206.-

Popova, G.V.(1980): On the practical use of reliability criteria in manual and automatic abstracting, in: Viks, U.(Ed.): Computational linguistics and related topics. Summaries of a symposium, Reval 24. - 26.11. 1980, pp.91-93.-

Schank, R.C.; Kolodner, J.L.; DeLong, G.(1981): Conceptual information retrieval, in: Information retrieval research, ed. by S.E.Robertson; C.J. van Rijsbergen; P.W.Williams. London et al.; Butterworths, pp.94-116.-

Skolnik, H.(1979): Historical development of abstracting. Journal of Chemical Information and Computer Science 19,4, pp.215-218.-

Strong, S.M.(1973): An algorithm for generating structural surrogates of English text. Computer and Information Science Research Center, The Ohio State University, Columbus, Ohio, April 1973 (OSU-CISRC-TR-73-3).-

Taylor, S.L.; Krulee, G.K.(1977): Experiments with an automatic abstracting system. Proceedings of the ASIS. Vol.14. White Plains, N.Y.: Knowledge Industry.-

Wellisch, H.H.(1980): Indexing and abstracting. An international bibliography. Santa Barbara, Calif., US: ABC-Clio, 308pp.-

TEXT UNDERSTANDING AS A KNOWLEDGE-BASED PROCESS

Claus-Rainer Rollinger
Hans-Jochen Schneider

Technische Universität Berlin
Institut f. Angewandte Informatik
PROJEKT KIT, Sekr. FR 5-8
Franklinstr. 28/29
D-1000 Berlin 10
Fed. Rep. of Germany

ABSTRACT

The task of "understanding texts with computers" can be divided into two main problem areas: first, the representation of the meaning of a text in a formal semantic representation language that is rich enough to express such complex meanings, and secondly, the interpretation of these semantic representation structures - e.g. to summarize a text or to answer questions about the text. The paper discusses the ways in which these processes (representation and interpretation) are knowledge based and argues that domain-specific knowledge is necessary to extract implicit information. In order to make these processes efficient, special interests in a text (reasons for reading it) have to be taken into account. This paper outlines some consequences of the assumption that "understanding" is an action influenced by goals and interests of the acting subject.

1. Introduction

The task of "text understanding with computers", as it is seen from the point of view of Artificial Intelligence, can be divided into two main problem areas: first, the meaning of a text has to be expressed in a formal language - a semantic representation language (SRL for example), i.e. the surface structure of the text (the natural language utterances) has to be transformed into expressions of this formal language. This translation process is knowledge based, as we will show in 2.1. Secondly, these semantic representation structures are interpreted in order to summarize the text (from a certain point of view), to answer questions about the text, or to extract information from it. This interpretation process has to be knowledge based, too, as we point out in 2.2. Section 3 outlines some consequences of the assumption that "understanding" is an action influenced by goals and interests of the acting subject. It is based on the observation that special interests in a text (reasons for reading it) have to be taken into account in order to make further progress in text understanding. The interest in a domain is distinguished from the interest in a new text that the subject wants to understand and assimilate. The latter, however, is influenced by the

former, by certain external conditions, and by the quality of the text itself. Understanding is knowledge based, but the depth of understanding depends not only on the extent of previous knowledge but also on the interest in the text.

2. Aspects of the SRL formalism

The semantic representation language SRL, still under development at the project KIT in Berlin, is a propositional, many-sorted formal language that is based on the notation of Kalish/Montague's description of symbolic languages for formal theories (Kalish/Montague 1964). We distinguish between terms and formulas: Only formulas can be true or false, or (in the context of uncertain reasoning) (Rollinger 1983a) there can only be evidence for formulas and not for terms. Terms are names and descriptions of objects (Habel 1982), so they have no truth or evidence value, even though there can be an evidence value for the formula expressing "This object has the description xyz". The most important entities of the language are the operators, which produce terms or formulas. Each operator has a type TYPE; t means that it is a term maker and f that it is a formula maker; m is the number of variables bound by the operator, n the number of terms, and p the number of formulas that the operator operates with.

$$\text{TYPE (operator)} := \langle t/f, m, n, p \rangle ; m, n, p \in \mathbb{N}$$

For all objects in a domain, there are sorts and the condition that one object belongs to exactly one sort but is compatible with more than one. The sorts are ordered in a lattice (Habel/Reddig/Rollinger 1981) where the arcs

correspond to the ISA-relation (Quillian 1968, Brachman 1979). With the sorts and the TYPE of an operator we construct an argument-type mask (ATM) where the sortal conditions of the arguments are defined. Finally we define the case frame of the operator (similar to Fillmore 1968), even for complex causal concepts like "because" and "but". With these concepts we are able to define optional and obligatory cases as arguments of the operator, and we are able to determine which argument (a term, or a formula representing an event) is under focus (Rollinger 1983b).

2.1. The representation structures of texts

In our approach to text understanding we distinguish three different structures which have to be built up in order to generate the semantic representation of a text. The argumentative, causal, and logical relations among the propositions (actions, events, states) of a text are represented in the argumentative structure. The names and descriptions of the objects, the relations between these objects, and the relations between the objects and the events (states and actions) they are involved in are represented in the referential net (Habel 1982). For each external object an internal representative is generated. The life-times of objects as well as of events are both represented with points of time in the temporal structure. There the relations between life-times and their relation to the "official" time axis (Wunderlich 1970) is represented. These three structures enable us to make the coherence of the text explicit (Grimes 1980) since the causal, referential, and temporal chaining is represented.

Über Schweden wurde der Tiefdruckwirbel Juno schwächer (pl), so dass an seiner Südfanke die Zufuhr der feuchtkühlen Luft nach Deutschland

(p2) am Donnerstag nachliess (p3). Gleichzeitig breitet sich das Hoch Alfons nordostwaerts aus (p4), waehrend sein Kern von Frankreich zur Schweiz zog (p5). Daher wurde in Norddeutschland die Schauerbildung weitgehend unterdrueckt (p6).

For this part of a weather report from July 2nd, 1982 (Tagesspiegel, Berlin) we generate the following argumentative structure:

```

p1 ---causes--> p2 ---modifies--> p3
  ^
  |
  | happen
  | at the same
  | time
  v
p4 ---causes--> p6 ---inferred--> p7
  ^
  |
  | happen
  | during
  |
  p5

```

p7 = The weather was pleasant on Thursday.

This is the argumentative structure built up by the system. It is the result of processing p1 through p6. On the surface we have the chain:

p1, so dass p2 (mod. p3). Gleichzeitig p4, waehrend p5. Daher p6.

If "so dass", "gleichzeitig", and "daher" are defined as relations among propositions, then one problem is to find out the propositions that are related to each other. An inadequate syntactic solution could be to directly link the propositions that are connected by these concepts. This would lead to the following structure:

```

so-dass ( p1,p2 ) & gleichzeitig ( p2,p4 ) &
waehrend ( p4,p5 ) & daher ( p5,p6 )

```

To achieve an adequate structure, knowledge about the events (states and actions) described by the propositions has to be taken into consideration. Only if we know that France is near Germany and that a high-pressure area

is responsible for pleasant weather, can we connect p4 and p6 by a causal relation. After having processed these propositions, the system has built up (beside the argumentative structure) the following knowledge:

```

p1 : < focus, + >
p2 : < focus, +- >
p3 : < focus, - >
p4 : < focus, + >
p5 : < focus, +- >
p6 : < focus, + >
p7 : < focus, + >

p1 : < tense, past >
p2 : < tense, past >
p3 : < tense, past >
p4 : < tense, past >
p5 : < tense, past >
p6 : < tense, past >
p7 : < tense, past >

p1 : < reference, [p2] >
p2 : < reference, [p1, p3] >
p3 : < reference, [p2, p6, p7] >
p4 : < reference, [p3] >
p5 : < reference, [p4] >
p6 : < reference, [p3, p7] >
p7 : < reference, [p3, p6] >

```

This information is necessary in order to process the next proposition as well as in order to interpret the argumentative structure. It should be obvious that the process of understanding implicit relations among propositions makes extensive use of general knowledge about the events, states, and actions typical for the domain in question. The algorithm for building up the argumentative structure is described in more detail in Rollinger (1983b).

As an example of a referential net we introduce the referential object (refo) r5 as the internal representative of the low-pressure area "Juno".

r5 SORT : low-pressure area
 NAME : "Juno"
 DESKR : iota (x , OBJECT-OF (x , p1))
 DESKR : eta (x , ABSCHWAECHEN (x , r4))
 DESKR : iota (x , LOCATION-OF (p2 ,
 SUEDFLANKE (x)))

The refo r5 has the sort "low-pressure area" and the name "Juno". We can refer to r5 with the (internal) descriptions "the object of event p1", "the object that weakens at r4" (with r4 as the location "above Sweden"), and "the object, the south of which is the location where event p2 took place". As we see, the refo's are linked with each other by the descriptions, and they are linked with the propositions in which they are objects, agents, or locations, and so on. Much geographical knowledge is needed in order to understand that the location of the object r5 is not only "above Sweden" but also "above Scandinavia", and that Scandinavia is also influenced by this low-pressure area.

Generating a new refo implies finding all of its properties that are not given explicitly in the text. With this information it becomes possible to recognize an object as one that was introduced earlier in the text. The referential net, attributes of refo's, and the problem of under- and overdetermination of refos is described in more detail in Habel (1981 and 1982).

The temporal structure is concerned with the representation of the life-time of the events (states and actions) and their relations to the real time axis. Therefore, the main entities are intervals (of time) and points (of time). The resulting time-objects as the basis of the temporal structure are linked to the propositions. A time-object is characterized in the following way: It is marked whether the object stands for the life-time of an event or

for an interval of the real time axis; depending on a point of the real time axis, the position of the object is marked, e.g. past, present, future, present and future, and so on; the third property marked the time-object is whether it is a point or an interval. These objects are related to each other by the relations "equal", "before", "after", "overlaps", and so on. A text could be understood completely if all (life-)time objects of the events could be connected (directly or indirectly) with the real time axis, or (if there is no possibility of linking the time objects to this axis, as is the case in fairy tales) if the resulting net is a single coherent graph. In Guenther/Habel/Rollinger (1983) the temporal structure is discussed with respect to inferential processes that enable the system to close the gaps of this structure.

In order to link time objects in an adequate manner, real world knowledge is necessary: the typical duration of events (states and actions) has to be known in order to understand the connective "waehrend" (during). Also, the typical course of (causally related) events is important information used in the understanding process.

2.2. The Interpretation of SRL structures

In this section we will give two examples of how to interpret the resulting representation structures described in 2.1. Imagine someone interested only in the weather of the last few days (for "interest", see section 3). For this person, the following summary of the weather report would be generated:

The low-pressure area Jung weakened above Sweden, and the high-pressure area Alfons above France spread northeastward. Therefore, the weather was pleasant in northern Germany.

This summary of the whole weather report is generated according to the information on the degree of focussing the propositions. We distinguish three grades of focus: - stands for very little focus, + stands for medium focus and + for high focus. The valuation of the propositions with respect to focus together with the time structure is the basis for selecting the relevant propositions for the summary. The referential objects have to be described in such a way that the reader of the summary is able to build up descriptions that are complete with respect to location and time. Therefore, the location of Alfons (in the past) has to be part of the summary. The argumentative structure of the propositions of the summary is the result of the complete argumentative structure of the text. The result of this interpretive process is the basis for a generation component which is able to verbalize the propositions. We have not yet implemented this part of our system.

After receiving the summary, the question "How was the weather in southern Germany?" would be a reasonable one. In order to answer this question the system has to activate its inferential (real world) knowledge, because no explicit information about the weather in southern Germany is available. Using the following rules, the system is able to generate the answer "I suppose that it was pleasant".

Rule 1: If the center of a pressure area is above a country, then the weather of the countries that are in the neighborhood of that country is similar to the weather there.

Rule 2: A high-pressure area in the summer implies dry and warm air.

Rule 3: Dry and warm air implies pleasant weather.

Fact 1: Southern Germany is in the neighborhood of Switzerland.

With this knowledge, the answer can be inferred. An important aspect of

1

this process is the ability of the underlying inference machine to process uncertain inference rules like rule 1 and to produce uncertain "facts" (Rollinger 1983a and Emde/Schmiedel 1983).

3. Attentiveness and interestingness in text understanding

In the investigation of human understanding two relevant aspects have to be considered: 1) Human beings can be regarded as goal-oriented acting systems. Although it seems that humans do not always act in a goal-oriented manner, for our purpose we are interested especially in actions instigated by goals. 2) Human beings can be seen as information-processing systems. The reading and understanding of texts is a process based on the computation of information as received by (visual and/or acoustic) perception channels.

As we know, information processing is not an accidental event but a goal-oriented, purposive action; therefore "understanding" has to be considered as a purposive action, too. This leads to an interesting and important task: to analyze the relationships between a) the object that has to be understood, b) the action 'understand', and c) the goals of the subject determining its acting. In his paper 'Interestingness' Schank (1979b) does not consider the active aspects. The inadequacy of his approach results from not regarding 'understanding' as an action.

If the 'direction of acting' (Lewin 1942) (the concrete actions processed in a special situation) is directly influenced by certain motivations and therefore determined by 'fields of interest' (Gust/Habel/Rollinger 1981) and 'goals', then an explanation is needed for the dependencies between the

action 'understanding' itself (in the context of text-understanding systems) and those factors. Their influence is clearly reflected by the attentiveness with which the action is processed, and different degrees of attentiveness seem to express the varying interest for the object that has to be understood.

In focussing on the action 'understanding', an important relation between the knowledge of a system and the action itself emerges: the description of events in a special domain (a text) is understood (by a system or by a human being) on the basis of existing knowledge. Extensive knowledge allows for deep understanding, whereas limited and badly structured knowledge of the domain in question will be a hinderance. Although profound understanding will always be limited by previous knowledge, we cannot assume that any text in that domain is understood in the greatest possible depth. The extent to which a text is processed depends on its importance as reflected in goals and interests rather than on the amount of knowledge in a domain.

The well-known text-understanding systems FRUMP (DeJong 1979) and SAM (Schank/Abelson 1977) are knowledge based: FRUMP uses sketchy scripts, whereas SAM uses very detailed scripts; the specificity of scripts obviously represents the extent of knowledge of the system. One consequence at Yale, however, is that the extent of knowledge of a system determines directly the amount of processing: the more is known, the more there is to process. Therefore FRUMP simulates a human being who knows only a little about a domain and is very limited in understanding more, even when he tries. This system can only skim a text and will never be able to understand it completely. SAM, on the other hand, has no choice but to understand a text fully and is forced by its profound knowledge.

The inadequacy of this approach is obvious: a human being has the ability to skim a text even if his background knowledge is superb. To formulate this argument more precisely: only knowledge that is extensive enough allows for successful skimming of the text. Consequently, FRUMP is not able to simulate true skimming but simply ignorance. This behavior can also be observed in the systems of the "new generation" at Yale (e.g. IPP (Lebowitz 1980) and MOPTRANS (Lytinen/Schank 1982)).

One must realize that the mental functions basic to the process of understanding (namely, a) the degree to which knowledge structures are "filled up" or specified, b) the reorganization of knowledge structures (induced by generalizations, for example) and c) other processes triggered by the new information) do not depend on knowledge structures alone but also are influenced by fields of interest and the goals. They appear to be responsible for the quality of the filling up of knowledge structures: the precision, subtlety, differentiation, and completeness with which the information will be stored and also to what extent a knowledge structure is connected and tied in with a net of other such structures.

Selective perception - that is, deliberate or unintended separation of relevant (and interesting) information from irrelevant - is the fundamental prerequisite for the ability to skim and is definitely human. It seems superfluous to enlarge upon this ability that enables people to process the huge amount of information they involuntarily receive all the time, but it is just this phenomenon that a theory of understanding must adequately describe and explain.

The Schank/Abelson (1977) approach, using scripts, is based on the

observation that people act easily in well-known situations (they have a lot of scripts in that domain), whereas they seem restricted in unknown settings (they still have to assemble the rules of this world). In the end, however, the Yale systems show a strangely opposite result: their extensive knowledge about a domain only slows them down instead of immediately putting them in command of the situation. This is the result of neglecting selective perception as a paramount function to be explained. Such an explanation seems possible if we regard the selectional process as a focussing on essentials (topics of main interest at a certain point in time) and if we accept that the process - the partial filling of knowledge structures - is greatly influenced by interest. The existence of complex knowledge structures like scripts is still essential for economical use of storage space, but selective perception solves the problem of making economical use of computing capacity and time.

Only if we assume 1) that the amount of information being processed remains constant at any point in time and moreover 2) that the density of information given is constant throughout the text, may the speed of reading be taken as an expression of attentiveness. These idealizations will be refuted soon, but they are helpful in demonstrating the effect of varying attentiveness. Besides the terms 'density of information', 'interestingness' and 'reading speed', the theoretical terms 'amount of information in a text', 'amount of information understood by the reader', 'knowledge of the reader', and 'energy for extracting information' have to be discussed in order to explain 'attentiveness' which - as we must stress - is quite different from concentration. Attentiveness is a rather quantitative aspect, whereas concentration qualitatively influences the processing of information. The fast reading of a specific text would necessarily be done with little

attentiveness (to that part of a text), whereas reading it (very) slowly would call for high attentiveness.

Attentiveness depends not only on interestingness: factors like limitations of time and space directly influence attentiveness (Marcus 1980). Expectation failures (described in Schank 1979a and Schank 1982) and results of checking new information against the knowledge of the system are other factors: receiving only well-known and consistent information will slow attentiveness down (the system is bored). Perceiving information that is quite new and that improves the consistency and the coherence of the system will enhance attentiveness; however, attentiveness will also increase if the information is inconsistent with previous knowledge, but in this case it is increased in order to resolve the contradictory situation. If the contradictions cannot be resolved or if it is decided to regard the new information as being wrong and/or the next information produces even more contradictions, then the attentiveness will slow down again.

If a special task has to be executed (e.g. summarizing a weather report) the attentiveness of the reader will be directed to special properties of the text - the attentiveness is clearly influenced by the reader's interest in executing the tasks. The producer of a text has several possibilities to manipulate the attentiveness of the reader: paragraphs, chapters, emphases, elements of style. We do not discuss these factors in this paper, but in order to simulate a human being reading a text, they must also be taken into consideration.

To summarize, we claim that it is not the interest in a whole domain that is expressed by attentiveness. Attentiveness rather depicts the interest in

the specific 'object' that has to be understood. This interest may change dynamically and quickly, whereas the interest in the "home" domain of the object in question is more static and does not change that dramatically.

4. Summary

In section 2 we gave a rough characterization of the semantic representation language SRL and its main features. In 2.1 we pointed out that the bases of the representation structure of a text are the argumentative, the referential, and the temporal substructures. If they can be built up out of a natural language text, reflecting its coherence, then the first part of understanding a text is solved. In 2.2 we show in what way the resulting representation structures can be interpreted in order to generate summaries or to answer questions about the text. These processes have to be seen as the second part of text understanding. That both processes depend on knowledge about the domain in question (world knowledge) was demonstrated with examples. In section 3 we pointed out that understanding not only depends on the domain-specific knowledge but also on the interest in the text and its information.

REFERENCES

- Brachman, R. (1979) : "On the epistemological status of semantic networks"; in: Findler (ed.) : "Associative Networks"; Academic Press, New York.
- DeJong, G. (1979) : *Skimming stories in real time*. Yale Comp. Science Dept. Research Report 136. New Haven, Ct.
- Ende, W. / Schmiedel, A. (1983) : "Aspekte der Verarbeitung unsicheren Wissens"; KII-Report 6, TU-Berlin, IAI.
- Fillmore, C.J. (1968) : "The case for case"; in: Bach/Harns (eds.) : "Universals in linguistic theory", London.
- Grimes, J. (1980) : "Context structure patterns"; Paper presented at the Nobel Symposium on Text Processing. To appear in the Proceedings edited by S. Allen.
- Guenther, S. / Habel, Chr. / Rollinger, C.-R. (1983) : "Ereignisnetze: Zeitnetze und Referentielle Netze"; erscheint in: Sonderheft Linguistische Berichte "Kuenstliche Intelligenz und Linguistik - Aktivitaeten in der BRD", Konstanz.
- Gust, H. / Habel, Ch. / Rollinger, C.-R. (1981) : *Language-World Systems : A communication model*. in: Models of dialogue : theory and application. Bolc/Hein/Haegglund (eds.). Hanser/McMillan : Munich / London.
- Habel, Chr. / Reddig, C. / Rollinger, C.-R. (1981) : "Term-Sortierung in MSRL"; SNP-Report 1/80, TU-Berlin, IAI.
- Habel, Chr. (1981) : "Zur Geschichte von Referenzobjekten - Teil I: Diskursmodelle". ms. Berlin.
- Habel, Ch. (1982) : "Referential nets with attributes" . in: J. Horecky (ed.) : *Proceedings of the COLING82*. North-Holland: Amsterdam.
- Kalish, D. / Montague, R. (1964) : "Logic - Techniques of formal reasoning"; Harcourt, Brace&World, New York.
- Lebowitz, M. (1980) : *Generalization and Memory in an Integrated Understanding System*. Yale AI Project Report No. 186.
- Lewin, K. (1942) : *Feldtheorie und Lernen*. in: Lewin, K. : *Feldtheorie in den Sozialwissenschaften*. Bern, Huber 1963.
- Lytinen, S. / Schank, R. (1982) : *Representation and Translation*. Yale AI Project Report No. 234.
- Marcus, M. (1980) : *A Theorie of Syntactic Recognition for Natural Language*. MIT Press.
- Quillian, M.R. (1968) : "Semantic Memory"; in: Minsky (ed.) : "Semantic Information Processing"; MIT-Press, Cambridge, Mass.
- Rollinger, C.-R. (1983a) : "How to represent evidence - aspects of uncertain

- reasoning" in: Proceedings of the IJCAI-83, Karlsruhe.
- Rollinger, C.-K. (1983b) : "Textverstehen und Textrepräsentation in SRL";
erscheint in: Rothkegel/Sandig/Weisgerber (eds.) "Text-Textworten-
semantik".
- Schank, R. (1979a) : Reminding and Memory Organization: An Introduction to
LOPS . Yale Univ. Dept.Comp.Science. Research Report 170.
- Schank, R. (1979b) : "Interestingness: controlling inferences". Artificial
Intelligence 12.
- Schank, R. (1982) : Dynamic Memory. Cambridge University Press.
- Schank, R. / Abelson, R. (1977) : Scripts, plans, goals and understanding.
Hillsdale, New Jersey.
- Sunderlich, D. (1970) : "Tempus und Zeitreferenz im Deutschen". Huber,
München.

HEURISTIC TEXT PARSING IN "TOPIC":
METHODOLOGICAL TOOLS FOR KNOWLEDGE-BASED TEXT CONDENSATION SYSTEMS

Udo Hahn / Ulrich Reimer

Universitaet Konstanz
Informationswissenschaft
Projekt Informationsvermittlung/TOPIC
Postfach 5560
D-7750 Konstanz 1

Abstract

The methodological tools of TOPIC, a knowledge-based system for text condensation and information management, are outlined with emphasis on its text parsing devices which are specifically tuned to identify textual cohesion and textual coherence data. These are essential to appropriate management of the knowledge acquired from the text. The text parsing system uses a word expert approach. In its current stage, TOPIC is intended to provide condensates on the informational level of indicative abstracts.

1. Knowledge-Based Natural Language Processing in Information Systems

As principles of knowledge-based natural language processing are beginning to influence the development of information systems substantially [01], their application to the task-domain of text processing information systems already has stimulated research on

- the access to data and text files through natural language dialogues
- the acquisition of new knowledge items from analyzed texts and storage in a knowledge base (automatic synthesis of data bases)
- abstracting of text according to different levels of explicitness
- paraphrasing text according to different user models

- relating the contents of separate texts (involving the time factor, subject matter of these texts)
- the generation of text from a knowledge base
- different modes of access to the knowledge extracted from texts (verbal mode vs. graphical mode)
- the translation of (sublanguage) texts
- intelligent text-editing and text-critiquing

Systems realizing various of these options for 'textual information management' are currently under way, some of which put special emphasis on the condensation (abstracting) of expository texts [02]. To illustrate this point, we describe TOPIC, a system for textual information management and condensation of expository texts, currently under development at the information science department at the University of Constance, West Germany.

2. TOPIC:

Text-Oriented Procedures for Information management and Condensation of expository texts

TOPIC is a system for text information management transforming expository full texts into appropriate textual or graphical condensates (abstracts, summaries, synopses, etc.), and supplying interactive graphical tools for abstract-tuned information retrieval [03].

TOPIC as described below is intended as a research programme which can only be realized in several steps. The current state of TOPIC corresponds to its first step - when necessary, remarks are made to distinguish future perspectives from current activities.

2.1 System Requirements

There are at least 3 basic requirements posed to the TOPIC system:

(i) Analyzing full texts

Expository full texts are to be processed by TOPIC, i.e. German language texts (approximately 2.000 - 4.000 words) dealing with product information and product tests in the subject domain of information and communication technologies for office automation

(ii) Generating condensates (abstracting)

These texts have to be reduced to less complex condensates, still supplying the main topic(s) of the original text while leaving away many of the subtle details of it (abstracting)

(iii) Text Information Management

TOPIC has to take into account user requirements to allow cooperative and effective interaction with its knowledge base providing access to the knowledge which has been acquired during the course of text analysis according to

- * different modes of interaction: verbal vs. graphical
- * different levels of specificity/complexity of system output
- * different kinds of informational target units: facts, extracts, condensates, etc.

2.2 System Design Principles

According to these requirements some basic modelling decisions were taken, effecting the system as a whole:

Decisions reflecting the analysis of full texts:

- (i') Within the domain of question-answering systems and natural-language interfaces to data bases highly specific parsing techniques and strategies have been elaborated, particularly suited for corresponding tasks (handling of interactive dialogue sequences, procedures for deep understanding of small quantities of phrases, etc.).

In the same way, independent procedures have to be developed for the task-domain of text processing.

Evidence supporting our view against far reaching structural equivalences between the processing of question-answering dialogues and lengthy texts was gathered from cognitive science, where computer simulations of reading behavior have been run which plausibly demonstrate human's integrative processing of texts in terms of the availability of schematic knowledge for immediate information processing, variability of time to process the single words of a text, and the immediacy principle of text-level integration (organization of information according to text structure) [04].

Therefore, TOPIC relies upon a knowledge-based approach to text processing applying to a high degree text-specific parsing mechanisms particularly well suited to the analysis of expository texts for the purpose of text condensation.

- (i'') The analysis of lengthy texts has to take into account chaining phenomena above the sentence level which serve the texts' organization on the level of textual cohesion and coherence.

Text linguistics has gathered evidencing data, that there are various structural means in language to make up a text:

Textual cohesion is illustrated by intra and inter sentence level connective phenomena like anaphora/kataphora, conjunction, elliptical expressions, lexical cohesion due to semantic relations, etc. [05]. A cohesive part of a text is here referred to as text constituent.

Textual coherence is the property of texts to connect cohesive parts of the text (text constituents) according to

- conceptual schemata (e.g., frame knowledge base) underlying text processing
- several linguistic macro devices: expanding, shifting, splitting of text constituents in terms of theme/rheme, topic/focus [06], or speech act patterns
- formal means of text structuring: paragraphs, typography, etc.

Furthermore, cognitive psychology has provided challenging methods to describe this over-all organization of coherent texts using story grammars [07].

Generating condensates is a research problem that can be split into several subtasks according to different informational levels of condensed texts [08]. The current version of TOPIC, actually, is restricted to the simplest and most commonly spread type of abstracts, i.e. indicative abstracts, with fairly simple means of providing graphical output.

Finally, this approach leads to the following considerations:

- (ii') Contrary to the requirements of deep understanding of natural language expressions in question-answering and special applications of text

understanding systems, indicative abstracting has to be based on strategies which systematically concentrate on the analysis of the major concepts of the text, thus achieving only its shallow understanding; this corresponds to the provision of the main topic(s) of a full text by standard indicative abstracts.

The problem of concept-centered text analysis has been extensively tackled in the domain of information retrieval leading to well-known approaches to the description of texts via key words/phrases (indexing) [09]

Surely, one has to augment the methodological tools for text analysis to account for text propositions or high-level speech acts (evaluative assertions, recommendations, etc.) when more sophisticated levels of abstracting are to be realized. However, in the case of indicative abstracting these requirements can be deferred at the current stage of the project.

- (ii'') As condensation of texts is primarily a knowledge-based process with varying degrees of linguistic guidance TOPIC's knowledge sources are organized in a way, that its (semantic) world knowledge component is superior to its (syntactic) linguistic knowledge.

Representation of world-knowledge is based on a frame model [10] which is characterized by the following properties:

- * simple concepts are structured by the relation between a frame header and its associated slots
- * single isolated concepts are connected by semantic relations
- * assigning slot entries to slots of a frame (slot filling) provides efficient tools for additional structures and relations of concepts

Representation of linguistic knowledge is based on a word expert model [11] which is characterized as follows:

Word experts are 'lexical grammars', i.e. each word of a natural language has attached to it a decision net to identify varying readings according to different contexts. The context of each lexical item is mainly tested by conditions on

- * states of the knowledge base: the existence of semantic relations and structural properties among concepts
- * states of the text: occurrence of lexical items in certain positions in the text (contiguity relations)

The representation techniques for world knowledge and linguistic knowledge are in perfect agreement to the concept-oriented approach applied to indicative abstracting. On the other hand, this approach seems to be extendable to support even deeper levels of text understanding as is necessary for higher options of text condensation (informative and evaluative abstracting).

Text Information Management principles lead to the follow suggestions:

(iii') We analytically separate the text parsing procedure, i.e. the structural analysis of a full text leading to its mapping on text representation structures, from the text transformation procedure which maps a text representation into the representation of condensates [12]. This allows for the flexibility needed to provide informationally differing levels and modes of output.

(iii'') Operating on a purely representational level, text transformation procedures have to be developed which distinguish 2 perspectives:

- * condensation transformations have to generate condensates differing in specificity according to various types of abstracts, i.e. indicative, informative, and evaluative abstracts (these transformations presume the availability of representation and parsing mechanisms for the most sophisticated condensation type when lower level condensates are to be generated)
- * visualization transformations have to generate different formats of output:
 - * textual mode: natural language abstracts, significant full text extracts
 - * graphical mode: zooming and navigation options in text graphs and table graphs for fact retrieval

2.3 System Organization

TOPIC's overall organization is based on the distinction of a text parsing component consisting of a word expert system operating on a frame knowledge base and a text transformation component.

In the following, the TOPIC project will be described with emphasis on its text parsing devices [13].

TOPIC, essentially, is a knowledge-based system with focus on semantic parsing, i.e. incoming text will be mapped directly onto some sort of 'semantic' knowledge structures.

Basically, the kind of semantic knowledge structures used in TOPIC are an extended version of the frame representation technique. Formal syntactic and semantic conditions lead to a frame data model [14] which is realized by

TOPIC's knowledge base management system. It does all the frame/slot activation, slot filling, activation and weighting interpretations primarily needed to achieve an adequate conceptual representation of the contents of the texts.

TOPIC's text parsing procedure is primarily determined by the occurrence of significant terms in the text under analysis. The basic activation pattern is as follows: each text token that can be morphologically matched to an item of the world knowledge gets instantly activated in the knowledge base or its activation weight gets incremented if it is already activated. In the course of analysis weighted conceptual activation patterns are evolving which exhibit a certain degree of connectivity (activation clusters). As an example, imagine the following sequence of incoming lexical items:

... microcomputer ... printer ... Z-80A ... HP-125 ... RAM ...

Readers disposing of specific knowledge in the domain of information technology, in effect, know much more than is simply indicated by this set of terms, as they exhibit a certain degree of structural organization and connection due to the world knowledge in this domain. Now, what is modelled in TOPIC's knowledge base, among other things, is this standard structural organization of hierarchically or associatively ordered concepts common to the specific application domain. Graphically this might be expressed as follows (the graphic on the following page is in essence a display of the parse net, i.e. the resultant structure of the text parsing procedures in the world knowledge):


$$\frac{51}{10}$$

This characterizes the state of the knowledge base after having processed the text fragment below:

Grundsätzlich unterscheidet sich der HP 125 von ähnlichen Personalcomputern am Markt dadurch, daß er Terminal und Rechner in sich vereinigt. Auf der einen Seite steht das Terminal, das ueber eine Z80A-CPU sowie ueber 32-KByte-ROM und 16-KByte-RAM verfuegt; ein eigenes Mikroprozessorsystem, das Tastatur, Bildschirm und Schnittstellen steuert. Auf der anderen Seite enthaelt das Gehaeuse des HP 125 einen Systemprozessor, ebenfalls eine Z80A, die ueber 64-KByte-RAM verfuegt und unter CP/M laeuft.

TOPIC's knowledge base always reflects the current state, according to the text's world description. By now, we have provided only limited means to achieve this requirement, as we assumed, so far, that a 1:1 correspondence must hold between terms of the text and items of the knowledge base. Clearly, this will be a too simple assumption.

This is the point when linguistic knowledge comes into play. We will discuss basic contributions of linguistic knowledge to the parsing process in terms of textual cohesion and textual coherence data.

Consider the following text fragment featuring anaphora, a basic tool to achieve textual cohesion:

... provided by this micro. Nevertheless, this computer ...

According to the activation patterns of the knowledge base, the occurrence of 'computer' will stimulate the activation of the equivalent concept in the knowledge base. This will be definitely wrong, as the author of the text intended to refer to 'micro' again. Only stylistic reasons caused him to

select 'computer' as a candidate pro-form; one might also have chosen 'it', etc.

To rearrange the knowledge base according to the author's intentions, we formulate a first requirement to an appropriate parsing procedure:

- (1) recovery from illegal temporal activation states in the knowledge base due to linguistic constraints

As a further simplistic assumption of the pattern-activation cycles of the knowledge base, these mechanisms are perfectly unsensible towards effects owing to the contiguity of proximate lexical items in the text also used to enhance textual cohesion, e.g. demonstrated by:

... supply of a disk drive is strongly recommended for the APPLE-II ...

The knowledge base has available structural knowledge concerning the association between both terms, but this knowledge is by no means sufficient to support constructive actions, e.g. slot-filling; as there are a lot of linguistic constraints to account for, the parser has to provide sufficient indication to stimulate or to preclude constructive actions in the knowledge base, as in the case of the inclusive contiguity relation between 'disk drive' and 'APPLE-II' above. This is provided by a special class of word experts, so called 'frame experts', which account for regular (generally predictable) constructive operations on the knowledge base. They also handle the phenomenon of anaphora, as stated above, as it is a regular pattern of textuality, too.

A quite opposite effect of contiguity is illustrated by the phrase below:

... HP-125 is supplied with a disk drive as against the APPLE-II ...

Again, the parser has to provide explicit indication to stimulate or to preclude constructive actions in the knowledge base, as in the case of the adverse contiguity relation between 'disk drive' and 'HP-125' resp. 'disk drive' and 'APPLE-II' above (due to 'as against'). This is provided by individual word experts, each for a word, to capture its regularities as well as idiosyncracies which account for irregular (generally unpredictable) constructive operations on the knowledge base. In the example above one has to provide word experts for 'as' and 'against'.

We state as second requirement to an appropriate parsing procedure:

- (2) exhaustive provision of legal activation states of the knowledge base due to linguistic contiguity relations (inclusive, adverse, etc.)

Finally, the parser has to provide tools to capture text linguistic phenomena on the macro level of text organization, i.e. to indicate text coherence supplied by

- * purely formal means of text structuring as indicated by paragraph, chapter organization of texts
- * linguistic means to indicate textual macro organization (e.g. constant themes, splitting rhemes, linear progression in texts [06])

We state as third requirement to an appropriate parsing procedure:

- (3) provision of indicators for cohesive segments of text (= text constituents)

In general, the task of the parsing system is to neutralize systematically interferences which would otherwise set the the knowledge base to illegal (1) or deficient (2) states and so to (re)arrange the knowledge base to conform completely to the actual contents of the texts independent of the linguistic

expressions used. Furthermore, cohesive parts of texts are identified according to the principles of text coherence data (3).

There are, in principle, several methodological tools to achieve such knowledge management by parsing mechanisms [15]. However, we decided on an word expert approach [16] to partial text parsing [17] because of

- * its modelling of the dominance of semantics over syntax
- * its assumption that most of natural language's information content lies in the individual word sense, rather than in any central or uniform system of rules
- * a homogeneous coupling of concept based representation of world knowledge and word based linguistic knowledge
- * variable degrees of specificity of analysis according to different levels of correspondence between a text and the knowledge base applied
- * the ability only to access relevant terms while skipping the remainder of a text

TOPIC's word expert system is in an early stage of experimental implementation on an ONYX C8002 micro-computer under UNIX. The system will be programmed in C. For communication, contact the authors of this paper.

NOTES

- [01] Cf. WALKER 1981, DeJONG 1983.
- [02] Text condensation systems, primarily concerned with abstracting of expository texts are described by FUM et al. 1982, HAHN et al. 1982, HOBBS et al. 1982, TAYLOR/KRULKEE 1977.
- [03] TOPIC's graphical component will be developed as independent man-machine interface under TOPOGRAPHIC (HEESE 1983).
- [04] The simulation system READER testing these assumptions is described in THIBADEAU et al. 1982.
- [05] For details see HALLIDAY/HASAN 1976.

- [06] A sketch of linguistic strategies to achieve textual coherence in terms of theme/rheme/focus devices is provided by DANES 1974.
- [07] Story grammars are originally due to RUMELHART 1975.
- [08] By different informational levels of condensed texts we refer to different types of abstracts, i.e. indicative, informative, and evaluative abstracts.
- [09] Cf., e.g. a non-knowledge-based approach to full text indexing in ROSTEK 1979.
- [10] The frame concept is originally due to MINSKY 1975. The remarks below are already tuned to the design of TOPIC, but they draw heavily on the original conception of frames.
- [11] The word expert concept is originally due to RIEGER 1978. The remarks below are already tuned to the design of TOPIC, but they draw heavily on the original conception of word experts.
- [12] Currently, we are not planning to implement a generation component for condensate texts. Instead, we attempt to experiment with different graphical modes of output. For a discussion of the advantages of graphical man-machine interaction, see HEESE 1983.
- [13] Suggestions how to model TOPIC's text transformation component are given in some detail in HAHN et al. 1982 and mentioned here only insofar as the text parsing is concerned.
- [14] An elaborated technical description of TOPIC's frame data model is given in REIMER/HAHN 1983.
- [15] WINOGRAD 1983, pp.361-410 examines the most commonly spread architectures of parsing systems: augmented phrase structure grammars, transformational grammars, chart rewrite systems, augmented transition networks, pattern matching systems, situation-action parsers.
- [16] To quote the original and basic ideas of the word expert approach: Word expert systems are a general approach to natural language meaning-based parsing in which the unit of linguistic knowledge is the word rather than the rewrite rule. In the word expert parser, knowledge about language is distributed across a population of procedural experts, each representing a word of a language, and each an expert at diagnosing that word's intended usage in context. Word experts ask questions and exchange information in coming to collective agreement on their word's meaning; see RIEGER/SMALL 1981, for a more comprehensive examination of the word expert approach, cf. SMALL 1980. A thorough description of the word expert system as adapted to TOPIC's requirements is provided in HAHN/REIMER 1983.
- [17] We strictly distinguish parsing techniques, e.g. word expert parsing as a special case of situation-action parsing (cf. [15]), and parsing strategies, e.g. top-down, bottom-up, islands, etc. Experiments applying text-specific parsing strategies to knowledge-based text understanding systems are currently exploring the effects of 'skimming'

(DeJONG 1979) in terms of integrated partial parser architectures (SCHANK et al. 1980). Non-knowledge based partial parsing of texts is illustrated in ROSTEK 1979. Combining both, is roughly the approach taken by TOPIC which we refer to as heuristic text parsing.

REFERENCES

- DANES, F. [1974]
Functional Sentence Perspective and the Organization of the Text. In: F. Danes (ed): Papers on Functional Sentence Perspective. The Hague, Paris: Mouton, 1974, pp.106-128.
- DeJONG, G.F. [1979]
Skimming Stories in Real Time: An Experiment in Integrated Understanding. Yale University, May 1979 (Ph.D.Diss.)
- DeJONG, G.F. [1983]
Artificial Intelligence Implications for Information Retrieval. In: Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. National Library of Medicine, Bethesda, Maryland, June 6-8, 1983, pp.10-17.
- FUM, D. / G. GUIDA / C. TASSO [1982]
Forward and Backward Reasoning in Automatic Abstracting. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.83-88.
- HAHN, U. / R. KUHLER / U. REIMER [1982]
Konzeption und Aufbau des automatischen Textkondensierungssystems TOPIC. Konstanz: Universitaet Konstanz, Informationswissenschaft, Projekt Informationsvermittlung/TOPIC, Sept. 1982 (= Bericht TOPIC-1/82).
- HAHN, U. / U. REIMER [1983]
Word Expert Parsing: An Approach to Text Parsing with Distributed Lexical Grammars. Konstanz: Universitaet Konstanz, Informationswissenschaft, Projekt Informationsvermittlung/TOPIC, Sept. 1983 (= Bericht TOPIC-6/83); (to be published in: Linguistische Berichte 1983).
- HALLIDAY, M.A.K. / R. HASAN [1976]
Cohesion in English. London: Longman, 1976 (= English Language Series 9).
- HEESE, C. [1983]
Die Konzeption einer interaktiven graphischen Mensch-Maschine-Schnittstelle fuer das System TOPIC: TOPOGRAPHIC. Konstanz: Universitaet Konstanz, Informationswissenschaft, Projekt TOPOGRAPHIC, August 1983 (= Bericht TOPIC-4/83).
- HOBBS, J.R. / D.E. WALKER / R.A. AMSLER [1982]
Natural Language Access to Structured Text. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.127-132.
- MINSKY, M. [1975]
A Framework for Representing Knowledge. In: P.H. Winston (ed): The Psychology of Computer Vision. New York 1975, pp.211-277.

- REIMER, U. / U. HAHN [1983]
A Formal Approach to the Semantics of a Frame Data Model: Extended Version. Konstanz: Universitaet Konstanz, Informationswissenschaft, Projekt Informationsvermittlung/TOPIC, Juli 1983 (= Bericht TOPIC-3/83). (A short version will be published in: Proceedings of the 8th International Joint Conference on Artificial Intelligence. 8-12 August 1983, Karlsruhe, W. Germany)
- RIEGER, C.J. [1978]
Computational Linguistics. In: W.O. Dingwall (ed): A Survey of Linguistic Science. Stamford/ Conn.: Greylock, 1978, 2nd ed., pp.97-134.
- RIEGER, C. / S. SMALL [1981]
Toward a Theory of Distributed Word Expert Natural Language Parsing. In: IEEE Transactions on Systems, Man, and Cybernetics SMC-11. 1981, pp.43-51.
- ROSTEK, L. [1979]
Methoden des partiellen Parsing fuer das automatische Indexing - Syntaxgraphen zur Analyse von Sprachmustern. In: R. Kuhlen (ed): Datenbasen - Datenbanken - Netzwerke. Praxis des Information Retrieval. Vol.1: Aufbau von Datenbasen. Muenchen etc.: Saur, 1979, pp.251-282.
- RUMELHART, D.E. [1975]
Notes on a Schema for Stories. In: D.G. Bobrow / A. Collins (eds): Representation and Understanding. New York: Academic Pr., 1975, pp.211-236.
- SCHANK, R.C. / M. LEBOWITZ / L. BIRNBAUM [1980]
An Integrated Understanding. In: American Journal of Computational Linguistics 6. 1980, pp.13-30.
- SMALL, S. [1980]
Word Expert Parsing: a Theory of Distributed Word-Based Natural Language Understanding. College Park/Maryland: Department of Computer Science, University of Maryland, Sept. 1980 (= TR-954).
- TAYLOR, S.L. / G.K. KRULEE [1977]
Experiments with an Automatic Abstracting System. In: Information Management in the 1980's. Proceedings of the 40th Annual Meeting of the ASIS. Vol 14. Chicago, Illinois, September 26 - October 1, 1977. Pt.2: Full Papers. White Plains/N.Y.: Knowledge Industry, 1977, A10-A14, B1-B6.
- THIBADEAU, R. / M.A. JUST / P.A. CARPENTER [1982]
A Model of the Time Course and Content of Reading. In: Cognitive Science 6. 1982, pp.157-203.
- WALKER, D.E. [1981]
The Organization and Use of Information: Contributions of Information Science, Computational Linguistics and Artificial Intelligence. In: Journal of the American Society for Information Science 32. 1981, pp.347-363.
- WINOGRAD, T. [1983]
Language as a Cognitive Process. Vol. 1: Syntax. Reading/MA etc.: Addison-Wesley, 1983.

CONSENSUS IN THE MEDICAL SCIENCES
AND THE IMPLICATIONS FOR INFORMATION SYSTEMS

Paper delivered to the Fifth International Research Forum in
Information Science (IRFIS 5) held in Heidelberg, FRG,
5 - 7 September 1983

by

Michael Brittain

Loughborough University, U.K.

Introduction: changing concepts of information provision

Most library and information services attempt to provide as much information as any user could conceivably require, as quickly as possible, and in a form most convenient for the user. For the most part information providers have assumed that the more information supplied the better, but other concepts are possible, and increasingly we realise that the other concepts may be preferable. It is common to talk about information paucity, but the idea of users and consumers being overloaded with information has attracted less attention.

The desire to produce an information rich environment has often concealed the very high cost of inefficient and wasteful information services, especially when they reach saturation point and subsequently overload users. The economics of information provision are beginning to be taken more seriously. It is now clear that a comprehensive library and information service, with high recall and with local backup document delivery service, will be impossible in the future in many situations. In the past such comprehensive services have been common in universities, research establishments and in some central government departments. However, the vast majority of practitioners in all fields have never been provided with such comprehensive services. Even in the medical sciences and healthcare fields, outside the teaching hospitals and medical research institutions, practitioners make do with relatively small libraries and information services.

Where restrictions have been placed upon the number of information services or the extent of a library collection attention has been paid to the concept of core collections of journals, and other documents. However, it is not always possible to identify a clear core of material for any particular user group, especially over relatively long periods of time. As a consequence collections in many libraries have grown

piecemeal, sometimes very much in a trial-and-error fashion, in an attempt to meet current local requirements.

There are political, economic, and scientific reasons for a change to more refined information services. Undoubtedly there is still a need for large comprehensive libraries of documents and data to provide a backup for the new generation of information services, which are less library and institution bound than in the past.

In an attempt to see the way in which more refined services can be developed in various fields of medicine I am currently exploring, with financial backing from the British Library Research and Development Department, three concepts of communication and information provision. They are: (1) consensus knowledge, (2) ignorance on the part of users, and (3) the cost (penalties) of ignorance. In this paper my examples will be drawn mainly from the medical sciences, but I suggest many of the ideas can be developed and applied in other disciplines and professions.

Alternative strategies and concepts

The new generation of information services will be more refined. By refined I mean that services will be more relevant to user needs; they will give users only what they can use/apply; they will provide data as well as bibliographical references; they will not provide bibliographical references as a rule of thumb when users have very little chance of making use of them; in many cases new services will provide consensus information; and, finally, they will take into account the costs of information provision as well as the costs of not having the information or having too much information.

The cost effectiveness of information still remains a problem. It is possible to specify the absolute costs of

a service, but the costs of having or not having the information are much more difficult to quantify. It is unlikely that much progress will be made in this direction. In the field of medicine there is overwhelming evidence that relevant information at the right time is of such great importance that it should be provided, even when the costs are high. But this does not argue for 'overkill' information services where users are provided with large numbers of bibliographical references in the hope that the user will pick up a few that are relevant. The costliness of such unrefined services has rarely been accounted. The 'overkill' approach is particularly prevalent in fields where there is low consensus, because neither information providers nor users can define with accuracy what they require. It is often assumed that this is characteristic of the so called soft disciplines like geography, sociology, economics. But the situation applies in some aspects of nearly all fields (with the possible exception of the physical sciences) and is certainly a problem in the medical sciences, as will be seen from the examples given below. The 'overkill' approach should not be confused with high recall. Sometimes researchers demand high recall, even if the relevance of most of the retrieved information is low, because the penalties of missing (say) one piece of relevant information are high. But this situation can happen only in areas with high consensus. The high costs of high recall can be justified. Outside such special situations high recall is often the order of the day: but to little effect. Indeed, it is often counter productive. The high costs involved cannot be justified. Such information services are wasteful. It is this problem I wish to consider in this paper.

In the present context we refer to a new generation of information services, but these should not be confused with data banks, data archives, or online retrieval services. These may indeed form part of the new generation of information services, but what we are concerned with at the present are

knowledge information systems, based upon consensus. They may contain data retrieval functions, even conventional bibliographical retrieval services, but they give much more weight to problem solving and question-and-answer services. Already prototype and operational services exist. Some are mentioned below.

Although librarians and information scientists play an important role in the design and operation of the new generation of information services, such services cannot be initiated and operated solely by librarians and information scientists. Some of the reasons for this are given below. They relate mainly to the problem of obtaining consensus among groups of experts about what knowledge should be in an information system.

Major factors involved in the design of knowledge based information systems

I have suggested three major factors must be considered: consensus knowledge, ignorance, and cost penalties of ignorance.

We now need to consider each of these factors, the way in which they interact with each other, the way in which they can be assessed, and the role they play in information services. It is noted that each factor is a dimension; that is, we can have varying degrees of consensus, ignorance, and penalty. Consensus is the most important factor. Without a knowledge of the degree of consensus it is impossible to attribute ignorance in any precise way, and by implication we cannot assess the cost or the penalty of not knowing. Some examples will make this clear. Three examples from the field of medicine are considered now. The first is the diagnosis and treatment of exotic diseases in the United Kingdom. This is an area where there is a high degree of consensus among the experts. The second example is from hepatology where there is an inter-

mediate degree of consensus. The third concerns treatment regimes for cancer, where there is a low degree of consensus.

First let us deal with exotic diseases. These are defined as diseases imported from warm climate countries, usually by infected people, but occasionally by animals or infected material. They include *Falciparum Malaria*, *Filariasis*, *Schistosomiasis*, *Tuberculosis*, *Hepatitis B*, *Dysentery*, *Cholera*, *Rabies*, and *Leprosy*. The incidence of exotic disease in the U.K. has increased during the last decade, for a number of fairly obvious reasons.

There is fairly high consensus among tropical medicine specialists about the diagnosis and management of exotic diseases (Haddock, 1984). In other words the specialist knows what *Malaria* or *Leprosy* is, how to diagnose it, and how best to treat it. However, for various reasons that I cannot go into in this paper, there is a great problem of communication within the medical profession in general, and between the medical profession and the general public about the identification, diagnosis, and treatment of exotic diseases in the United Kingdom. Within the specialist community there is relatively little difficulty in deciding what information should be put into textbooks, data bases, specialist information services, and so forth. This is because there is a high degree of consensus among the experts - even if the majority of medical practitioners do not know about this information.

It is easy to attribute ignorance in the case of mis-diagnosis and mis-management of exotic diseases - although it should be noted that there is still some disagreement about the efficacy of certain drugs in the treatment of some exotic diseases. The penalties of ignorance are great. The general practitioner in a small rural community who does not diagnose *Rabies* in an infected person places the patient, the community, himself, and indeed, the medical profession and the Local Health Authority at great risk.

The second example concerns research into hepatitis and its diagnosis and treatment. Hepatology is a field where the scientific and medical literatures are fragmented, overlapping, and to some extent redundant. Few papers are of the quality that make them long-lived. It therefore takes much time and effort to find important information and assess its relevance. Surprising as it may seem in a field such as medicine, the practice of 'weighing' the evidence is often seen, as I have noted it is in many of the social sciences (Brittain, 1979). John Carson (1983, p.5) notes that "Judgement on the validity of data is often swayed by the number of articles retrieved rather than their quality. What is needed is a process of purification. This would be particularly welcome ... where doctors need quick answers to patients' problems; preferred drugs, dosage and duration of treatments, and adverse drug reactions ...". Conventional bibliographical retrieval services are incapable of providing such results.

Consensus in hepatology is variable. There is a good deal of agreement about the identification of the hepatitis virus, and the different types. But it is a field characterised by information richness rather than information paucity. The dedicated researcher and expert do of necessity seek, obtain, sift, and judge a wide range of documents and statements about hepatitis. The majority of practitioners are not able to do this, because of limited time, and because of other interests, motivations, and inclinations.

The new generation of information systems - those that are knowledge based - involve the critical judgement by experts of new data and concepts as soon as they are published, and their integration into a structure so that their location can be easily identified. One such system is the Hepatitis Knowledge Base developed by the National Library of Medicine. In its construction one of the first tasks was to find a method of developing consensus amongst a group of experts about the

data, information, paragraphs, sentences, and statements that were put into the base. The resulting consensus ranged from complete agreement to differing views. Where conflict exists the evidence is given and referenced. The Hepatitis Knowledge Base has been expensive to construct and continues to be expensive to up-date and maintain. In part this is because it is expensive to obtain consensus judgements. If this field had a high degree of consensus the information contained in publications could be put straight into the knowledge base.

The third example concerns treatment regimes in cancer. This is an area in which there is low consensus amongst researchers and physicians. The causes of cancer are still not fully understood. This is an example of scientific ignorance, of a most basic nature. For the most part we are not dealing here with this type of ignorance. We are dealing with the ignorance of users of theories, concepts, and facts that are already known. The treatment of cancer, especially chemotherapy, differs widely, one country to another, and also one institution to another. Oncologists I have spoken with have pointed out the low degree of agreement about treatment regimes and have gone as far as condemning the so-called 'cowboy' approach to treatment. This may take place when a physician visits another country, observes that a particular drug is being used in the treatment of cancer, and without scientific evidence returns home and tries out the new drug, perhaps in very different circumstances.

In a field like cancer, it will be particularly difficult to create a knowledge base, at least with respect to treatment regimes. There is a low degree of consensus and therefore it is difficult to identify and quantify ignorance. For example, if a physician in London does not know of the use of a particular drug for the treatment of a particular type of cancer in (say) New York, he will not be regarded as ignorant, for a variety of reasons. His local colleagues may also have no knowledge

of the New York practice, which may not have been communicated outside a local group of physicians and researchers. Even if the New York practice has reached the learned journals it will be competing with many other such claims and there is little pressure for other cancer specialists to take much notice - because there are so many claims to the discovery of a wonder drug that subsequently turn out to be false and quickly pass into the archives. His colleague in tropical medicine unaware of the symptoms of Rabies, and therefore mis-diagnosing a case of Rabies, would indeed be regarded as ignorant and reprimanded. Likewise, his colleague in anaesthetics who did not know an accepted procedure, or a well-known risk, could be called ignorant. It follows of course that the consequences and the penalties are much different. At one level the penalties may be all the same. The patient dies. The Rabies victim dies; the patient undergoing surgery dies during anaesthetics; and the cancer patient dies. But the role of information in these tragedies is very different. In the case of cancer there is no failure to communicate. We just do not yet know how to treat cancer. In the other two cases there is a communication failure; perhaps during training, perhaps due to faulty information services. Also the physician himself may be careless or negligent, and so forth.

The role of information providers

If the factors of consensus, ignorance, and cost/penalty are so important in the development of information services, we may ask why information specialists have not been much concerned with these factors. There is one very clear answer. Information specialists (even those responsible for the generation of mechanised and on-line retrieval services) have for the most part been concerned only with bibliographical reference services. Question-and-answer services (like the Leigh University LEADERMART system), data banks, and the more recent knowledge-based services are rare. Scientists themselves

have not been overly concerned with knowledge-based systems, partly because the scientific communication system is itself, viewed globally, a consensus developing machine. As Ziman (1979) has pointed out scientific knowledge is public. It is made visible for all who want to see, replicate, and so forth. In the physical sciences of physics and chemistry the communication system produces consensus quickly and usually without fault. This is done by publication, presentation of results of research at conferences, the invisible colleges, and so forth. There are always disagreements at the frontiers of research in science, but such disagreements are usually short lived and resolved (Collins, 1981; Koester, Sullivan and White, 1982). For the most part each new generation of students learn and experience within well established paradigms, and only occasionally are there major scientific upheavals and revolutions.

Later in the history of science the biological and medical sciences developed, and at the research level they followed time-honoured communication activities of the successful physical sciences. Later still many of the social sciences followed similar paths, and it was expected that as they became more scientific a greater degree of agreement would develop about theories, concepts, and practices, and that subsequently they would be successful in solving social, political, economic, and personal problems. None of these disciplines or professions have been much concerned with the process of consensus development. Perhaps it was assumed implicitly that consensus could look after itself, as it appears to have done in chemistry and physics. But it has not. Many fields outside physics and chemistry face enormous problems at the level of professional communication, and in the development of information services which enhance their research and practice. The problems are not ones that can always be tackled solely from outside the profession or discipline.

A small group of scientists within each of the major disciplines are now interested in information problems. Their expert subject knowledge, combined with the special skills and technologies of the information scientists, are on the verge of producing a new generation of information services. Progress will be slow and costly. Scientific discovery and progress is itself slow, painstaking, and costly. There is no reason to suppose that we can telescope the processes of research and development found in science, in the field of information and communication. But the rewards will be great. For example, the NLM Hepatitis Knowledge Base was constructed in the first instance to provide an information service for viral hepatitis for those researching in the field, but there have been many spin-offs. Services are now available for clinicians, educationalists, nurses, and other paramedical staff. Such services are very different from the traditional bibliographical retrieval service where all users delve into the pool of bibliographical references and where the only difference in the service for (say) nurses and researchers was perhaps that the researchers retrieved (or were supplied with) more references than the nurses. In knowledge based systems, the questions put to the service will differ one group from another, and responses will be different at a qualitative as well as quantitative level.

References

- BRITTAİN, J.M. Information services and the structure of knowledge in the social sciences. International Social Science Journal, 1979, 31(4), 712-728.
- CARSON, J. The Hepatitis Knowledge Base: a new direction in information retrieval. Report to the British Library Research and Development Department, February 1983. Report No. 5768 (unpublished).
- HADDOCK, D.R.W. Information surrounding the diagnosis and treatment of exotic diseases in the U.K. In J.M. Brittain (Ed.) Consensus in the medical sciences and the implication for knowledge based information services. Report to BLRDD, 1984 (in preparation).
- COLLINS, H.M. Son of seven sexes: the social destruction of a physical phenomenon. Social Studies of Science, 1981, 11(1), 33-62.
- KOESTER, D., SULLIVAN, D. and WHITE, D.H. Theory selection in particle physics: a quantitative case study of the evolution of weak-electromagnetic unification theory. Social Studies of Science, 1982, 12, 73-100.
- ZIMAN, J. Public knowledge. London 1979.

The Declining Readability of Scientific Papers as a Factor

Limiting the Transfer and Use of Knowledge.

R.T. Bottle

Department of Information Science
The City University
Northampton Square
London EC1V 0HB
U.K.

The older scientific papers, especially those published in the first two decades of this century, strike one immediately as being much easier to read than today's papers, even when one is familiar with the subject matter. The layout seems more spacious, the page size smaller and the print larger. These format factors certainly affect readability, but the style of writing has by far the greatest effect.

Readability may be assessed in a number of ways. Though most methods have their origins in the assessment of children's texts, they are based on the premise that familiar words are easier to read than unfamiliar words and that abstract concepts are more difficult to understand than descriptive passages. The corollary is that familiar words are mostly short and unfamiliar words are generally polysyllabic. Furthermore abstract concepts tend to be expressed in polysyllables. Readability studies have been well reviewed by Klare (1). One of the simplest techniques to use is the Farr, Jenkins and Patterson simplification of the Flesch 'Reading Ease' formula (2). It is also the method most amenable to computerisation and, in this connection, it is interesting to note that word processors can be programmed to compute the readability of the typescripts they produce (3). Normally readability measurements come out as a numerical score which can then be referred to a qualitative scale with bands such as standard, fairly difficult, difficult, very difficult, or the reading

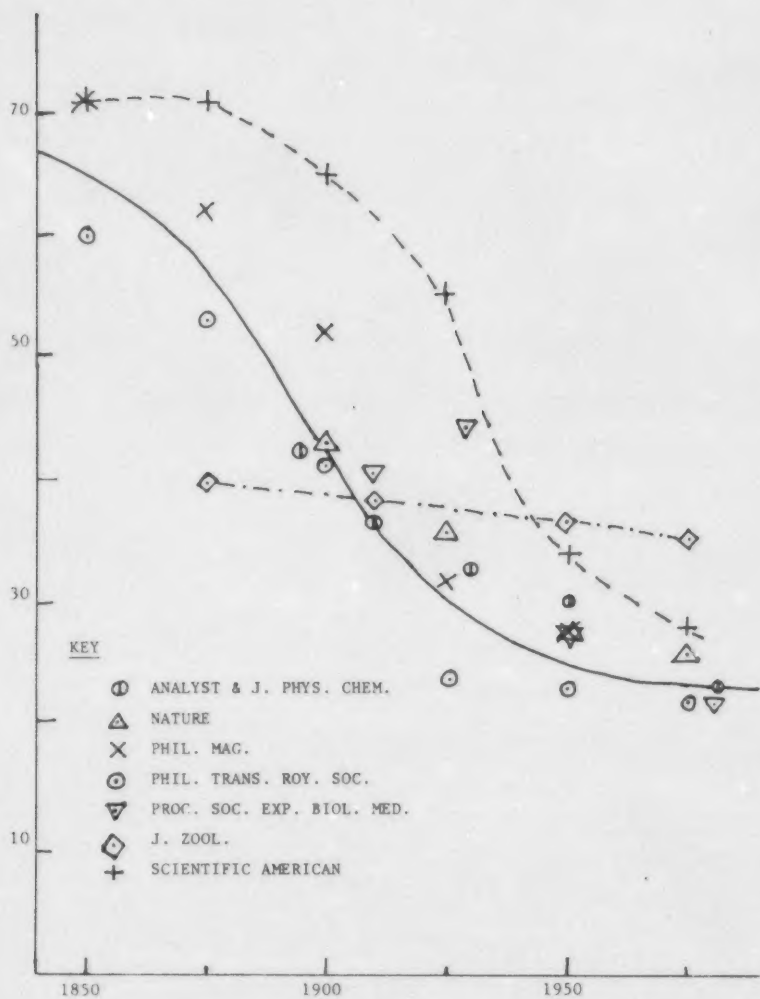
capability of school children, etc. I do not, however, claim that these methods, developed for children's literature, give absolute results when applied to scientific texts. I do, however, believe that they can usefully indicate relative changes in readability and indicate trends. For a useful discussion of the limitations of readability formulas applied to technical material, the reader is referred to six recent essays in IEEE Transactions on Professional Communication, especially (4). If the readability and format changes are examined over a period one can then see whether a gradual change occurred or whether there was a time when the change was particularly large.

Sardar (5) examined changes in readability over the period 1950 to the present day for four journals using the Farr, Jenkins and Patterson (2) simplification of the Flesch 'Reading Ease' formula

$$R = 1.6n - s - 31.5$$

where R is the Flesch 'reading ease' or readability score, n is number of single syllable words per 100 words and s is the average number of words per sentence. His data, redrawn in Fig. 1 show that four journals, Philosophical Transactions of the Royal Society, Philosophical Magazine, Nature and Scientific American, decline from readability scores of just over 70 classed as 'fairly easy' by Flesch (or 60 for Phil. Trans.) in 1850 to scores of between 65 to 40 ('standard' to 'difficult') in 1900, finally levelling out in the 25 to 30 range ('very difficult') in the third quarter of this century. The period of steepest decline in readability occurred between 1900 and 1930. It is interesting to note that Phil. Trans., a primary journal, led the decline, whilst Scientific American, a review journal aimed at the intelligent layman, lagged behind. Chemistry in Britain, however, maintained a constant score of 30 (borderline 'difficult/very difficult') during its first 11 years (1964-75).

FIG. 1
READABILITY OF SCIENTIFIC PAPERS



Similar Flesch reading ease scores were obtained by Goodsell (6) for medical papers published between 1970 and 1975. These were mainly on viral diseases and endocrinology. The scores ranged from 13 to 32 with a mean value of 23 ('very difficult'). It should, however, be borne in mind that the qualitative terms 'fairly easy', 'difficult', etc refer to the reading capabilities of American High School students and not those of scientists. Nevertheless these results point up a large relative decline in readability in this century.

Readability scores were subsequently determined for The Analyst, Journal of Physical Chemistry, Journal of Zoology and Proceedings of the Society for Experimental Biology and Medicine using the Farr, Jenkins and Patterson formula (2). Ten papers and 4 samples per selected paper were selected at random from volumes for 1980, 1950, 1930, 1910 and 1896. Numbers, other than single digit ones and abbreviations, were counted as single polysyllabic words. Single digit numbers were counted as monosyllables. Sentence lengths were also noted separately. Excepting the Journal of Zoology, the average readability for these primary journals was not significantly different than that for the common carrier journals studied by Sardar (5). The general shape of the curve was similar and the 1980 reading ease score was 23 ('very difficult'). (Journal of Zoology showed very little change in readability, falling from 40 in 1896 to only 35 in 1980. Presumably this is due to a somewhat different type of science being reported in Journal of Zoology than in the three more laboratory biased journals.)

There was little variation in sentence length during this century for these four primary journals. It tended to reach a maximum value around 1930 but this was only about 20% above the minimum. Whilst there was considerable variation in sentence length from 5 to 96 words in the 160 samples examined at each of the time intervals, the average sentence length in each paper showed much less

variation (from 10.8 to 38.5 words). The overall average sentence length was 23.0 words. This compares with the average value of 23.8 words given for Genre J (Scientific and Learned Writings) of the so called Brown Corpus (7) where the sentence lengths ranged from 1 to 116 words. The Brown Corpus analysed texts "first printed in 1961....in the United States". Genre J, however, seems to contain as much material from the social sciences and the humanities as from the natural sciences and technology.

It is to be hoped that the readability of scientific papers has now levelled out at Flesch's 'very difficult' level. At least the trend curve no longer points to papers becoming even less readable. (The very low scores reported below for patents and their abridgements show what could be achieved if we switched to 'patentese'!)

Currently we are investigating the readability of standards and pharmacopoeias. Preliminary results for British standards indicate that over the past 50 years there have been fluctuations in R values of between 65 and 11 (median = 42.3). There is, however, no definite trend. Ministry of Defense (MQAD) standards fluctuated between R values of 66 and 19 (median = 37.1) and also showed no definite trend (8). Studies on the British Pharmacopoeia (BP), British Pharmaceutical Codex (BPC) and Martindale's Extra Pharmacopoeia (MEP) showed that BPC fluctuated between R values of 36 and 27 in the period 1907 to 1978 but that MEP and BP declined appreciably in readability. BP fell from R=35 in 1967 to R = 16 in 1980 and MEP fell from R = 44 in 1887 to R = 17 in 1982 (9). A French speaking student is just starting a parallel study of the readability of scientific literature in French over the last 100 years and will also look at papers by individual authors and the changes over the course of their lifetime. I assume that frequent occurrence of compound words prevents analogous studies on German scientific texts.

Several other studies have determined readabilities at just one point in time. Pharmaceutical patents granted under the 1977 Patents Act were compared with similar ones granted under the earlier Patents Act. The average readability of the specifications were not significantly different at $R = 18.5$ and $R = 14$, neither were their author abstracts at $R = 2.5$ significantly different from the Patent Office abridgements produced under the earlier legislation with $R = 1.6$ (10).

Company financial reports on Ceefax and Oracle showed obvious similarities with the corresponding company press releases. Readability scores were therefore found to be highly variable. The mean readability of papers in an electronic journal (the BLEND Project) was found to be about 30 with specific passages in the range 53 to 16. Even information science and librarianship journals have been studied by a former colleague (11).

Ultimately authors are responsible for the readability of papers which they write. Advice on writing papers abounds. Particularly good advice is given in a saying attributed to Churchill "Old words are best, but old words when short are best of all". Figs. 2,3 and 4 illustrate such advice.

Fig. 2

SCIENTISTS DESIROUS OF INCREASING
THE READABILITY OF THEIR WRITTEN
COMMUNICATIONS SHOULD SUPPRESS THEIR
INHERENT PREDILECTION FOR POLYSYLLABIC
TERMINOLOGY AND MINIMISE SENTENCE
LENGTH.

Reading Ease Score = 12

Fig. 3

SCIENTISTS WHO WANT TO MAKE THEIR
PAPERS EASIER TO READ SHOULD USE
SHORT WORDS IN SHORT SENTENCES.

Reading Ease Score = 76

Fig. 4

WANT YOUR PAPERS TO BE EASILY
READ?
USE SHORT WORDS IN SHORT SENTENCES.

Reading Ease Score = 89

Fig. 2 is written in pseudosociological gobbledegook and has a reading ease score of 12. Fig. 3 shows a dramatic increase in reading ease score to 76 simply by replacing long words with short ones, and only a 15% reduction in sentence length. This is a slightly higher reading ease score than Tolkien's The Lord of the Rings where scores between 65 to 75 were observed. The style of Fig. 4 may be thought to be too informal for a scientific paper. With a reading ease score of 89 it is almost at the level of a six year old's reading primer.

Does the decline in readability mean that more long words are being used and hence that average word length has been increasing? Data were obtained for the number of characters per word in 17 scientific journals for 1910 to 1970 (12). Since five of these journals were not English language ones (two were German, two French and one Swiss) the average number of characters per word for the twelve English language only journals in the sample was calculated.

(The twelve journals examined included two from the field of chemistry, two from mathematics, two from biology, two from biomedicine and one each from physical sciences, astronomy, psychology and engineering.) These showed less than 3% variation from the overall average of 6.27 characters per word over the whole period. Thus there does not seem to be a direct relationship between mean word length and readability changes.

Probably of more significance is the fact that readability is affected by format and layout factors such as print size and the proportion of the page occupied by print. Sardar (5) found for the journals which he studied that the proportion of white space and the print size decreased over the period studied. Thus readability further decreased over and above that caused by the increased complexity of the English. These changes in format and layout factors have doubtless been brought about by the exponential growth in the number of papers submitted for publication and by inflation also increasing the pressure on editors to reduce the costs of publication (13).

Readability factors are obviously approaching the stage where they are affecting communication among scientists. Because 60 or 70% of primary scientific papers are now in English, a majority of scientists will have to read material which is not in their mother tongue. Perhaps even more important is that an increasing proportion of English language material is being written by scientists for whom English is their second language. (In a sample taken from three agricultural abstracting journals for 1980, one third of the English language papers originated from countries where English is not the mother tongue (14). This proportion is believed to be even larger in some areas such as biochemistry. A more detailed analysis of the stylistic causes of the dramatic decline in readability shown in Fig. 1 should lead to guide lines to authors and editors for improving written communication. In the new fields of

videotext and electronic journal systems, the incorporation of systems similar to that described by Kincaid et al. (3) could lead to this being done semi-automatically.

References

1. G.R. Klare. Assessing Readability. Reading Research Quarterly, 1974, 10, (1), 62-102.
2. J.M. Farr, J.J. Jenkins and D.G. Patterson. Simplification of the Flesch Reading Ease Formula. Journal of Applied Psychology, 1951, 35, 333-7.
3. R. Coleman. Introductory Speech at CNA (UK) Seminar on the Future of Chemical Documentation, University of Exeter, Sept. 6, 1982; J.P. Kincaid, J.A. Asgard, J.W. O'Hara and L.K. Cottrell. Computer Readability Editing System. IEEE Transactions on Professional Communication, 1981, PC-24, 38-41.
4. J.C. Redish. Understanding the Limitations of Readability Formulas. IEEE Transactions on Professional Communication, 1981, PC-24, 46-8; D.L. Plung. Readability Formulas and Technical Communication-ibid, 52-4.
5. Z. Sardar. Noise in Written Communication. M.Sc. Thesis. The City University, 1975.
6. M. Goodsell. The Readability and Presentation of Medical Literature. M.Sc. Thesis. The City University, 1978.
7. H. Kucera and W.N. Francis. Computational Analysis of Present-Day American English. Brown University Press, Providence, R.I., 1967.
8. A.H.E. King. M.Sc. Thesis. The City University, 1983. In preparation.
9. J.D. Powell. M.Sc. Thesis. The City University, 1983. In preparation.
10. L. Blakeborough. The Information Content and Readability of Patent Abstracts and Abridgements. M.Sc. Thesis. The City University, 1980.
11. C. Oppenheim. The Reading Ease of Information Scientist. Journal of Librarianship, 1976, 8, 77-8.
12. S. Russ. The Growth of Scientific Literature, 1900 - 1970. M.Sc. Thesis. The City University, 1979.
13. R.T. Bottle, J.S. Rennie, S. Russ and Z. Sardar. Changing Patterns of Chemical Communication I: Some Effects of Growth. Journal of Information Science, 1983, in press.
14. A.S. Greenwood. Regionality of Agricultural Literature. M.Sc. Thesis. The City University, 1982.

Treatment of Societal Problems and Information Systems:

An expanded concept of "Information- and Communication-technology"

Marianne Buder,
Freie Universität Berlin,
Projekt INSTRAT;
Gerhard Vowe
Technische Hochschule Darmstadt,
Institut für Politikwissenschaft.

IRFIS.BV

0. Overview

The following remarks try to emphasize the relevance of information- and communication-technology for the treatment of societal problems - i.e. problems on a macro-level.

Therefore an enlarged concept of information- and communication-technology is necessary, which includes not only hardware and software, but also brainware and orgware - on the level of interaction of macro-actors. This is possible on the background of a broad concept of technology and it is necessary because of an obvious crisis of the societal information- and communication-system.

The brainware offers options for change in some dimensions, which will be developed in the course of the paper. In respect of these broad-frame changes a general concept is not available at present. To develop such a concept a lot of empirical research work must be done. The orgware can be structured also in different dimensions. The optimization of orgware as an answer to evident deficiencies can be summarized under the concept of information-agency.

G. Vowe is responsible for the first part, where the concept of brainware is developed, whereas M. Buder is responsible for the second part, in which she is describing the options for orgware. So, differences in language and method can be explained by the co-authorship; nevertheless, we hope the general context will become transparent.

1. Background

A societal problem can be defined as a credit-debit discrepancy, which is seen as relevant enough as to draw the attention of macro-actors, whereby none of these macro-actors can offer a procedure, which is acceptable for all the other macro-actors involved.

A conflictual process of arrangement leads out of this problematic situation; during this process each macro-actor tries to hold on his version of a possible solution and places his specific resources. Conflict and arrangement are socially regulated up to a defined level, whereby these norms can assume problematic features.

Problems (for instance the social distribution of work) and the interactions around the problems are becoming more and more complex. The notion of complexity refers to the number of elements and different relations which should be taken into account in the evaluation of proposals for problem-treatment. The increase of complexity can be measured in the dimensions of time, space, social tension and cognitive depth.

In this context we are especially interested in the part, which information and communication (I+C) are playing or ought to play. Which relevance do they have for the generation of consensus, for the organization of conflict, during the agreement-process on procedures and so on? Which relevance should they get, if you set the normative target, that social conflicts and their treatment should be cultivated?

We want to design a model for surmounting the distances in time, space, status, interest, speech, knowledge and rules between social macro-actors through symbolic interaction.

Our grounding thesis is, that the context of societal I+C generates growing deficiencies in respect of the complexity of problems; the existing system of formal and informal components of societal I+C is overloaded.

A friction-free performance doesn't seem possible; revision of grown concepts and procedures is necessary. There are many indicators for the different shortages: Biases in the societal diffusion of knowledge; waste of existing knowledge; violent interaction instead of symbolic interaction; compulsion to spectacular symbolic interaction; blocking up of the implementation of knowledge.

We want to investigate first this I+C-aspect of societal problem-treatment and then make some remarks, how to optimize the societal problem-treatment by reconstruction and innovation of I+C.

2. Strategies for the Development of Societal Information- and Communication-Systems

The overload of the societal communication-system should be countered by development of the technology of the communication-system. But this isn't yet a clear defined task, because "technology" includes different facets, which are compatible with different strategies.

"Technology" should be defined as the socially and culturally influenced means and procedures of problem-treatment.

Which means and procedures how and at which time will be developed, isn't only determined by the physical circumstances, but is - partially - debatable between social positions and depends on the cultural situation. Means and procedures can unfold their virtual power in different forms.

Technology is selected knowledge about how to obtain a given objective. These options depend on different social, cultural and physical criteria.

Usual and evident is the identification of technology and artefacts. But in artefacts the selected knowledge is only incorporated. It is only one of its different forms of existence. Technology as knowledge - knowing how, more and more on the basis of knowing-why - can appear in non quested skills and articulated experiences of persons or - as socio-technology - in social institutions, rules, organization-structures, law, procedures and so on.

Especially that selected knowledge is of interest and of relevance, which is part of the background of our everyday-life, upon which the non quested means and procedures are built on. It functions as the horizon of all our activities: The technology of perception and the overcoming of space and time. (e.g. walking, watching, feeling, timing, conceptualizing, encountering, interpreting signs and so on). These technologies are forming the frame of human action and are part of identity and cultural commonness.

Such technologies, which are used by the members of a given cultural community, can be conceptualized as "Culture-Technology".

For instance: The computer is an incorporation of the technology of binary digitalization, which you can find just as well in other ways as basic clas-

sification-principle in our perception or as a structural principle in law or architecture.

Between the different steps of knowledge there are dependencies and junctions: An incorporation in artefacts is only possible, if knowledge does exist about the procedure independent of the physical shape; if this knowledge is socially compatible to a specific level, that means, socially acceptable and imaginable; if it rests upon grounding cultural patterns.

These four steps can be described and changed with different instruments:

- Artefacts can be physically described and changed
- procedural-technology can be logically and psychologically described and changed
- socio-technology can be described and changed (by sociology and political Science)
- cultural-technology by anthropology and philosophy.

In order to differentiate on this background the narrow complex of Information- and Communication-Technology - that is the choosen knowledge about the reduction of uncertainty by symbolic interaction - we will employ the often used concept-quadrupel of hardware, software, orgware and brainware: We will use the concepts in a slightly different way.

"Hardware" means the physically measurable material and energetic components of an incorporation of an I+C-technology.

"Software" means the given arrangements for a target-related operation of the hardware, especially the programmes.

"Orgware" means the rules for structuring the given components in time, space and logic order.

"Brainware" means the mental treatment of the problems as a basis for further technological arrangements.

In a structural entity - like the operation of a computer with a design-task - all these aspects are always present; when a technology has been developed and the system-engineers have gained experience, the aspects can be treated seperately.

In the beginning of computer-development the hardware was the main task and

all other aspects were implicit in its development. Now the other aspects have emancipated themselves and are starting their own process of theorization and optimization. The discovery of new aspects has opened new resources for better performance of the whole system.

Through separation and definition of more aspects it is possible; to open the scope for optimization; for instance by the definition of socio-ware - the art of embedding the other aspects in the social field; of culture-ware - the art of combining I+C-technology and cultural background; or of time-ware - the art of timing as an aspect of I+C-technology.

If one has figured out these separate aspects, the engineer can work directly on a special aspect - optimization can be worked out with a set purpose.

Therefore it's not necessary, to work at the same time on all aspects.

Technology-based strategy, then, should be guided by the principle of opening the scope through development of different ways of action and of synchronizing the actions under the different aspects.

We, now, want to detach this classification of I+C-technology out of the narrow micro-context and to transfer it to the societal level.

Hard- and software can be looked upon as the technical frame-work of societal problem-treatment. The functional components, which are part of each information system on each level, are also Input/Output (for instance Home-TV, ballot-box), transmission (e.g. telecommunication network, local space), storage (e.g. libraries, "collective memory") and processing (e.g. selection-rules of press, decision-rules of bureaucracy).

This framework of artefacts and programmes contains a cognitive structure of the field; the brainware of the I+C-technology for societal problem treatment is incorporated in this frame and can be reconstructed and changed.

The hard- and software of societal I+C-technology contains also a certain organizational structure, which can be reconstructed and changed as orgware.

Now we are provided with a differentiated concept of I+C-technology for societal problem treatment and can combine this concept with the main thesis above: If there is really an overload of the societal I+C-system, one has different possibilities of answer:

An evident and obvious possibility is to make changes in the hard- and software of the societal I+C-system: One invests more resources in new technical

infrastructure, develops a new and more comfortable equipment. This is the variant, which works in a classical engineer-scientific thinking.

We assume, that this variant, which puts the emphasis on innovation in hard- and software, is a necessary, but not a sufficient condition, to increase the performance of the societal I+C-system.

Instead, we want to stress the relevance of innovation within the cognitive structure of the societal problem-treatment: changes of the brainware as an aspect of I+C-technology. Changes in the brainware, then don't mean, that another corpus of knowledge diffuses during the process of problem-treatment; changes in the brainware mean, that other cognitive conditions prevail, under which knowledge can diffuse.

The other strategy, we want to plead for, is to change the organizational structure of the existing components and resources. These developments are not necessarily bound to parallel changes in the hard- and software. So, this strategy concentrates on optimization of the orgware.

Hence, we propose other preferences to be dealt with: Org- and brainware in preference to hard- and software; there are separate resources hidden, to increase the performance of the whole system and - as another argument for our emphasis - because every change in hard- and software needs and causes changes in brain- and orgware.

3. Changes in Brainware

The goals of change in some different dimensions of brainware-sub-goals of "Societal-Problem-Treatment through I+C" - and the corresponding instruments - sub-instruments of "Brainware of the societal I+C-system" - are listed in Tab. 1.

Dimensions	Goals	Instruments
Material Dimension: Which material basis requires which cognitive framework?	Variety of opportunities of operation and mix of artefacts in problem-, actor- and structure-adequate ways	* Brain / External memories in different levels * Sensually and technically enlarged perceptions * Face to Face Communication and Telecommunication Tendence: Gain of distance with increasing level of organization-effort
Time-Dimension: Which time-structure requires which cognitive framework?	Increase in the flexibility of the temporal order of actors, components and resources	* Coordination of the single developments * Variation of the rhythm of the overall-development
Substantial Dimension: What is necessary to know something about what?	Making available the knowledge in problem-, actor- and structure-adequate ways	Storage and access to: * Problem-knowledge * Actor-knowledge * Structure-knowledge * Knowledge about knowledge
Social Dimension: How should the knowledge be distributed and how should it be confronted with each other?	Treatment-adequate relation of conflict and consensus	* Logic of argumentation * Modalities of access to debate * Rules of controversy
Reflexive Dimension: Which knowledge about the cognitive framework is necessary and how should it be used?	Legitimation and increasing efficiency of the development of the cognitive framework	* Rules of the development of the cognitive framework * Meta-Communication

Tab. 1: Dimensions, objectives and instruments of "Brainware" for the societal problem treatment

To clarify the synopsis in Tab. 1, we want to explain one of the dimensions - the social dimension.

The brainware for societal problem-treatment can be described in its social dimension by the distribution of knowledge between the problem-involved actors (this has been done by the proponents of the so-called knowledge-gap-hypothesis) and of the ways, in which these different knowledge-resources are be confronted.

The goal for brainware in this dimension should be the stabilization of a relation of conflict (confrontation) and consensus (understanding), which is adequate to the challenges of the problem and the possibilities of the actors.

The instruments of brainware in this situation are for instance the logic of argumentation and reasoning (binary or other), modalities of access (who is entitled to know something and permits anotherone to know something at which time?).

In this dimension and in the same way as in the other dimensions some deficiencies and possibilities of development occure. This is listed in Tab. 2; here you can find also the ongoing research projects in the different dimensions.

Dimension	Deficits	Tendencies of development	Projects
Material Dimension	* Lack of group-communication * Lack of relief in higher communication-spheres	Increase of performance in breadth (covering different audiences) and in depth (take-over of intellectual routine-tasks)	Enrichment of teleconferencing-systems by elements of artificial intelligence
Time Dimension	* Overload of the participation-budget * Asynchronicity (social bias of access to knowledge and of capability to react)	Enlargement of time and acceleration of access through division of labour	Exploration of time-budget used for societal participation (especially I+C-part)
Substantial Dimension	Overload of the knowledge access and processing-possibilities	Enlargement of the individually controlled knowledge-space through more knowing-that-knowledge (Options, structures, consequences, procedures)	Exploration of - the part of indicative knowledge in individual knowledge - the changing-rate in public agendas
Social Dimension	* Blockade through mutual exclusion of positions * Coercion of agreement through extra-communicative means instead of conflict and consensus in a communicative manner	Training and normalization of differentiated understanding by dialectical logic (processual and not binary) and by variable synthesis	Reconstruction of successful understanding-processes (e.g. commissions)
Reflexive Dimension	Utility-Ethics instead of Communicative-Ethics (Precept of agreement by understanding)	Operationalization of communicative ethics for difficult situations (Time-shortages, different interests, unknown circumstances)	Exploration of codified and everyday used elements of ethics of communication (standing order, client-expert-interaction)

Tab. 2: Main deficiencies of existing "Brainware", tendencies of possible and desirable developments and ongoing research-projects

To illustrate the synopsis we take the temporal dimension for an example: If one describes the temporal dimension of the brainware of societal problem treatment one can find some obvious deficiencies:

For instance a social synchronization in the knowledge-access and in the capability to react do not exist: Some people have earlier more and better knowledge about problems and possibilities of treatment and can use this knowledge in a better way than others. There is another deficiency one can see in the overload of the participation time budget: Nobody is able to participate in all these roles, in which it would be necessary to participate (as a parent, at job, as a political subject, as a component of an ecological system, as member of a cultural context). One tendency of development, which can be strengthened by operations, is the enlargement of disponsible participation-time through political structures, which combine divisions of participation-labour and new methods of control and surveillance and which differentiate for instance goal-decision and routine-decision.

In order to utilize these possibilities to a greater extent, we have started some explorations for instance, to measure through questioning and diary the part of time budget for participation, especially for I+C, at engaged persons (shop-steward, member of a civil action group, party-activist, parents-delegate in school and other). Out of their organization of everyday-life we hope, to generalize means and procedures, to enlarge the participation-budget and to develop practicable models of timing.

Also in other dimensions explorations have started, which should join together to an overall-picture and an overall strategy for developing brainware as an aspect of I+C-technology for societal problem treatment.

There is no unifying concept for all these different dimensions available at present; that won't be possible until the empirical studies have shown further results.

4. Orgware

The changes in orgware cannot be assessed in complete isolation from the other "wares". As a rule, they presuppose shifts in the brainware (which can be very far ahead in time) and often require new hardware and software (as well as other technologies) - or they react to them with the intellectual impulse for new possibilities and/or realities.

Information and communication are concretised in the so-called practice of information, in a narrower sense in what is today usually called "information and documentation".

The structures of information systems rest on intellectual and organizational constructions which imply the respective, time-dependent "philosophy" of not only information work but also in general of how to cope with problems and conflicts in a way that appears (socially) adequate. And this is done, for the most part, by sharing out and allocating responsibilities, a division of labor at the macro level, which means a breakdown and organization of an ever complexer world (society) in quantities that are simpler to grasp (e.g. in branches, professions, institutions, sciences, markets, functionaries, etc.).

At the same time, today's information systems are dependent on certain "traditions", i.e. the gradually greater (or reduced) understanding of what information is as the subject of professionalized, human work, what (social) role is played by information and documentation and what functions/tasks information systems can and should have in the overcoming of social and individual problems.

Thus, on the one hand the realization of brainware can be found in orgware, while, on the other hand, it develops dynamics of its own after its establishment, which in turn generates brainware that can lead to new creativity and/or innovation - but also to bureaucratisation.

4.1 Characteristics and Deficits of Today's Information Systems

Attempts to organize information start, for the most part, at a time when there is no longer an adequate possibility of imparting findings and knowledge by (direct) communication, due inter alia to the increasing quantity and complexity of knowledge (although from the technological point of view the development of transport and communication technologies could have made communication easier). As is generally known, the channeling of knowledge has been entrusted to formalized information institutions, which try to perform their tasks of storing and imparting the knowledge on data carriers (i.e. from bibliographies to data banks, above all with a literary list of sources of varying quality and/or intensity).

The respectively available (new) technologies (file cards, punched cards, microforms and reproforms, EDP) have been tested and used to cope with the growing volume and more diverse content of knowledge. In this connection, the possible uses were not of inconsiderable significance for the respective structuring of information and documentation and the understanding of their methodology.

The information systems display various structural dimensions, which derive, derive, inter alia, from

- a) the relationship regarding the ties with the financially responsible body, i.e. the principal. Here the orgware is mainly visible in the form of administration;
- b) the internal structure expressed in hierarchies, sections, responsibilities (input, output, information transfer, etc.);
- c) the immanent structure that rigidifies because of the way the realisation of methods (formalization, standardization, classification, systematization etc.) are related to the technical field, contents and objects;
- d) the relationship vis-à-vis the user, which is mainly realized in various information services and products.

A deepening of the structural dimensions leads to various intermeshing organizational levels and patterns, which in the end constitute the organizational structure and "functioning" of the systems.

Today's information reality is distinguished by the various developments of these structural dimensions.

- a) The role of information units and systems is fixed (as intermediary institutions between the producers and users of knowledge) with the establishment of information and documentation as an independent field of work; this is done, inter alia, by setting them off from other institutions in other fields of work (e.g. librarianship, publishing).
Institutional rigidification especially increased in the 70s due to centralization efforts (external to information and documentation); they were supported by the technological possibility of larger EDP memories (hardware) and corresponded to the general (state) trend of rationalising by amalgamating and/or interlinking institutions of the same type.
- b) With the professionalization of those working in the field of information and documentation there evolved job concepts that determine the tasks, responsibilities and role of information specialists; the internal job concept is continued by corresponding training courses.
- c) The very difficult, methodical accomplishment of the information task contains the danger of becoming a l'art pour l'art occupation; the ever more complex information systems (including the application of complicated technologies) can only be operated by information specialists any more.
- d) Partly because of the "image polishing" indulged in by scientific information specialists and partly because of historical links to the organization of intellectual work, information work is still pursued mostly for science and technology (with emphasis on research) and therefore directed toward a limited sector of demand.

The range of services and capacity of the information systems derive from the competence of the institution, characteristics of work done by the information specialists, efficiency of the applied technologies; as a rule, individual services are formalized services specific to the subject matter. Information systems are available to every "adult" user, but he must adapt to the systems.

Today, information systems marked by rigidified structures are apparently no longer in a position to guarantee a supply of information within the scope of coping with more general problems and conflicts. A few reasons for the deficits are given in the following:

- Due to their rigidification information systems are hardly able to react to changes in the surrounding field any more.
- Information systems have created an island or ghetto for themselves inside the world with its conflicts and problems. Occupation with fixed knowledge, which necessarily leads to formalization, classification and technicalization, does not square with the real world, but it tries to look as if it were related thereto in a rationally understandable fashion.
- Therefore, the transfer of knowledge via formalised information systems is achievable only to a limited degree.
- Access to the knowledge is not facilitated by the systems. People who can transfer information are required, not so much as experts on the contents but as authorities on the complicated systems. The information specialist is an authority on the partial mastering of information problems, and therefore a "custodian" of the (information) system.
- Apart from intelligence, information systems provide no further help for the mastering of problems (e.g. advice).
- Intelligence and advice services are today assigned to other branches and institutions and are rarely directly linked to information systems but are, at the most, "intermediate users" (or they construct their own systems).
- With the reduction of the mastering of problems to information problems, the systems distance themselves from the reality of the world, and/or the user's actual problems.
- It is hard to overcome the barriers faced by the user. He is confronted by a strong medium that permits him hardly any feedback. As a rule, the interaction with the system is not communication for the mastering of problems.
- The neutrality of knowledge processing and transfer contradicts the case-related viewpoint of the user.

The isolation of the information problem from the "goings-on of the world" or the reality of the actors (those with problems) is not likely to guarantee an up-to-date, adequate supply of knowledge or be accepted by those with problems. These systems lack the variety, reflexivity and concern to be more than a kind of back-up for the transfer of the problem- or case-related information proper.

4.2. The Information Agency as a Professionalized Service for the Supply of Knowledge

If importance is attached to information and communication as a means of coping with social and individual problems, organizational forms must be found which can adapt to real processes of problem-treatment and befit the options of those with problems. This would be necessarily lead to a de-institutionalization of functions - with regard to today's information systems - which would be realised decentrally for the most part and in which information institutions and users would equally participate in a continuous process of interaction.

A possible and conceivable, professionalized service for the supply of knowledge that meets this demand is to be described in the following in the form of an information-agency model.

The structural dimensions of the information agency (IA) are five-tiered:

- within the IA, i.e. as regards the organization of the work done by the IA itself
- actor-related, i.e. as regards the interaction with the actor (user, person with a problem)
- area-related, i.e. as regards the area (branch, problem field) for which the IA is technically responsible and competent
- related to the surrounding field, i.e. as regards the tendencies, decisions and development affecting the area to be covered
- cooperative, i.e. as regards the inclusion of the IA in a network of agencies or other competent or responsible institutions or experts.

The tasks of the IA does not first begin at the time when an actor requires information to cope with his problem but extends to the entire problem-solving process (to the extent that the actor permits the IA to participate therein).

The IA is defined by the functions which are necessary for this task and which are conceivable in various organizational forms:

- The actual supply of knowledge takes place by means of the back-up function, which necessitates links with the information system and knowledge

stores (data banks, hosts, experts, etc.), as well as by the establishment and operation of the IA's own storage facilities. It serves the active and passive supply of information.

- The communicative component, i.e. the communication between the actor and IA for the mastering of problems, is realized by the advice function.
- The problem-determination function is the beginning of the common coping with a problem (midwife). It is up to the IA, together with the actor, to determine the causes of the problem in order to be in a position to prepare corresponding solution strategies. This necessitates (spatial) mobility on the part of the IA.
- The actor is assessed in the diagnosis function. This is necessary in order to develop strategies which fit the actor's needs and which satisfy not only the objective problem but also the actor's mentality, willingness to take a risk and interests.
- The problem-articulation function (maieutics) presupposes methods that lead to the recognition of a problem even without the active help of the actor (e.g. by provocation). At the same time, it necessitates orientation to external developments and influences on the actor's (potential) field of problems in order to recognize the consequences for him in good time (prophylaxis).
- The pathfinder (or path-searcher, path-preparer) function (boyscout) implies not only the strategies for the mastering of problems but also the accompaniment in the mastering of problems. In this connection, the choice of strategy is left up to the actor. For the IA this means the flexibility to adjust to the strategy respectively chosen. Thus, the pathfinder function does not mean, recommending, delegating, referring but advising, accompanying, attending to.
- This model can be realised only if information agencies and other competent services are linked by a network. In this case, the matter can be passed on to a more competent institution of the agency function.
- Supply and demand are regulated in the market function; not only the actor and IA can orientate themselves to the same but also potential producers of knowledge. (The basis for subsidies and/or financial support would also be found here.)
- The services function covers the services and products the IA makes available for up-to-date and continuous and/or prophylactic (prognostic) infor-

mation. They can be realised with classical data carriers as well as with new communication technologies (BTX, video disc, etc.). For this purpose, new forms of editorial organization, in particular, are required.

- Finally, the diachrony function, which includes the self-control of the IA itself, is responsible for the functioning of such a model. Reflexivity and variety of the IA are concretised therein, and from the same it is possible to derive not only acceptance and effect research but also (in addition to the technical competence) the ability of the IA as a whole to function in its role as a service for the provision of knowledge and mastering of problems (including, finally, the question of its necessary existence).

Concrete activities, which are intermeshed and depend on each other and therefore require a certain organization, also derive from these functions. But here orgware does not mean the administration of knowledge or rigidification of internal structures, whose existence alone provides justification of functions in activities, work sequences, preliminary work, services which are necessary in order to deal with problem-treatment processes that vary in type, some of which also undergo internal changes.

This results in a number of categories that decisive for the orgware of the IA:

- The IA does not evolve from a collection of knowledge or data which are made available in certain cases of problems but evolves from a field of problems itself.

Therefore, the individual IA will not be found at the level of a social "principal" but more likely at a level of collectives, groups, communities of interests. Only the sum of all IAs will cover the overall social need for the transfer of information and a decentrally organized network of agencies for the provision of knowledge with which to master problems or also cope with conflicts.

- The degree and structures of the organization depend largely on the IA's tie-ins and link-ups. They can be independent or part of a self-contained organization/institution (firm, plant, etc.), belong to a branch (association, company) or also develop in the form of current (possibly limited to a certain period of time) communities of interests with a specific subject matter.

- This affiliation (not dependency) will determine its spatial (geographic, regional) effectiveness and competence (even with decentralised usability on the basis of new technologies).
- This affiliation will also determine the regular clientele who, although hardly homogenous, have in common acceptance of the IA as a competent agency for the provision to help in the common or individual mastering of problems.
- Acceptance can only be achieved if the IA displays concern, i.e. the role of neutrality will have to be given up in favour of identification with the actor's problems. It also presupposes that barriers and fears (personal, financial) are eliminated.
- In addition to individual help (case-related, client-/actor-related) the IA also makes it possible to inform groups and collectives via the body representing the common interests.
- In this case the IA can also become the body representing the interests of a collective (on the basis, at least, of certain "positions of strength" through the supply of knowledge but also through the provision of direct information to the outside world (interlinkage of information work and publicity, for instance)).

At last categories to be mentioned already result in expanded room to maneuver with regard to

- the actor
- the sector of knowledge production
- the sector of knowledge transfer
- the "decider" and/or deciding bodies (macro-actors).

In this connection the IA plays a more active than passive role in as much as it

- reacts to the problem (case) of the actor (with the background of its own state of information),
- informs the (actual or potential) actor on the basis of its own information/knowledge on the sector and/or surrounding field,
- informs other actors on the basis of concrete problem situations (without, in so doing, impairing the interests of its clients),
- informs "responsible" agencies (deciders) on the basis of concrete problem

situations.

- informs "deciders" on the basis of the knowledge of the developments and their consequences (thereby intervening in the decision-making process),
- informs producers of knowledge on the basis of problem situations (e.g. orders research/development),
- evaluates its own work.

Compared with today's information systems there is a new understanding of the roles played by the information agency and the "information workers" therein, which in part belongs more to the brainware, but also has consequences for the real work:

- The provision of knowledge and transfer of information cannot be limited solely to the solution of information problems but must assume its function within the social, individual or collective mastering of problems.
- This leads to the necessity to create professionalized bodies for this purpose, which are not rigidified "systems" but can attend case-related problems in a flexible way.
- In order to be effective in actually "user-related" fashion and gain the acceptance of the actors, document- and knowledge-related neutrality must be changed in favour of actor- and case-related concern.
- The structure of the information agency must not be "system-preserving" due to its complexity, but must instead, contain in itself the possibility of providing help for the purpose of self-help.
- A professionalized body for the supply of knowledge must also have the possibility of preventing problems by means of prophylactic information. In part, therefore, the time at which information work is done will have to be shifted forward in order to be socially interesting and effective.
- The information agency will have to cooperate with other bodies and institutions dealing with information and communication. This includes, in particular, the opening of communication forums (e.g. exhibitions, conferences, workshops), which can take place independent of the IA or be initiated by it.

There are a number of starting points for information agencies, in not only the classical information units (e.g. of an intercompany nature) but also, and precisely, outside of information and documentation and independent of

today's information systems (e.g. within the framework of management consulting / Information broker). So far all of them perform only part of the IA functions mentioned. But they are in indication of the fact that the information-agency model is not utopian but is based on concrete starting points of very scattered information reality which as a rule does not feel a part thereof (a consultant from the chamber of industry and commerce does not see himself as an information specialist any more than an employee of an agency for technology transfer does), and which must be brought together and coordinated.

In the further development of the new information and communication technologies there are good chances for the realization of a network of information agencies. However, the decision to make basic technologies available in such a way that decentralised information networks can result from the initiative of those with problems will be a social and/or political one. For the possibility of an "information trade" in which anyone can participate presupposes as a basic social condition the general use and availability of technologies, services and information.

Today's information systems have developed out of the information institutions which various consumers first organized as a kind of self-help. The improvements in hardware and software will increase their efficiency only in-substantially or relatively.

It will be much more important, and surely more difficult, to find new structures and organizational forms of information transfer (which will have to be realized with improved hardware and software) that suits the property of information and communication as a tool for coping with conflicts and problems. The model of an information agency has been introduced as one possibility.

May the brainware of correctly understood information and communication keep the orgware from leading in this case as well to the "mapping" and "transformation" of the very lively and therefore conflict-ridden world, which must be put in order if information problems are to be solved.

ON THE IMPACTS OF MODERN INFORMATION
TECHNOLOGY ON INFORMATION NEEDS AND SEEKING:
A FRAMEWORK

by

Kalervo Järvelin
University of Tampere,
Dept. of Mathematical Sciences,
P.O. Box 607, SF-33101 Tampere 10, Finland

and

Aatto J. Repo
Technical Research Centre of Finland,
Technical Information Service,
Vuorimiehentie 5, SF-02150 Espoo 15, Finland

ABSTRACT: The impacts the new mechanisms for transfer and use of knowledge have on specific work situations and problems are changing the information that is relevant to this work. Our main problem in this study is how, in principle, this impact will affect information needs and related phenomena. Our approach to the problem is based on conceptual analysis and restructuring of information need-related concepts, and on knowledge work augmentation as analysis framework. A conceptual structure for the phenomena related to information needs and seeking, and a taxonomy of modern mechanisms for the transfer and use of knowledge, are developed in the paper. Using this framework, the effects of computer conference systems, database systems and expert systems on information needs and seeking are discussed. Finally, we present some general hypotheses and compare our framework with the other ones.

1. BACKGROUND

The role of information and knowledge in society has never been so emphasized as right now. Much is talked about information business, information technology and sometimes even of information needs. Our sincere aim is to outline here the impacts on the people working with information and knowledge, especially on the information they need and use.

There is a tradition of well over thirty years of information need (or user) studies in information science (Crawford -78). To mention a few, the use of information services (e.g. libraries), the users of these services and their preferences, information dissemination and flow, their information seeking behavior and information use, are examples of the foci of individual studies in this field (Menzel -66, Kunz -77).

The conceptual and methodological problems, and the problem of usefulness or uselessness of the tradition, have been recognized in the literature (e.g. Brittain -75, Kunz -77, Wersig -73). Most of the studies have focused on library and information service institutions and on document needs and use within this context. Only relatively recently user studies have attempted to assess work situations and work problems in detail, and then to relate, or deduce, from these findings the information needed to successfully and efficiently carry out work tasks, and solve scientific, social, economic, political and practical problems (Brittain -82).

Modern information technology is rapidly changing the specific work situations through office automation, e.g. text processing, electronics records, and publishing, computer-based messaging via Local Area Networks, public data networks and various personal supporting systems (Järvelin & Repo -82). The separate technologies are being integrated and they will provide working environments and work stations for personal and

group use in many professional domains. This seems to be the direction, though there will be trouble e.g. with the problems of standardization and costs. The progressive application of information technology aims at enhancing professional work to achieve speedier and better solutions to the problems encountered (e.g. Engelbart -62).

The specific goals of this study are:

- 1) Developing a conceptual structure to cover the different phenomena that are related to information needs and seeking.
- 2) Developing a taxonomy of knowledge work support systems to analyze and describe the effects on information needs and seeking.
- 3) Presenting general hypothesis on these effects.
- 4) Comparing a framework to earlier ones.

We discuss the relations of knowledge work, knowledge, and the information business in chapter two. The types of knowledge relevant to problem solving in knowledge work are also outlined. In section three we develop the conceptual structure for information needs and seeking. In chapter four we develop a taxonomy of knowledge work support tools and discuss the relations of information needs/seeking and the tools in various classes of the taxonomy. Three examples: computer communication systems, database systems and expert systems are discussed more thoroughly. Chapter five contains conclusions and discussion.

2. KNOWLEDGE WORK AND THE WORLD OF INFORMATION

Information society or knowledge based society is one of the most discussed concepts of today (e.g. Toffler -81, Martin - 81), yet it has remained quite loosely defined. A common definition is related with the amount of people working with information and knowledge; it has been said that when over 50 % of the people do knowledge work, the society has shifted to the age of information society.

Information or knowledge work is professional work where the acquisition, production, storing, retrieval, representation, use, communication and creation of information plays a major role (Drucker -66), (Engelbart -75). The most, if not all, parts of professional problems and tasks of scientists, executives, engineers, journalists, teachers, social workers, lawyers etc. are knowledge work. These and the like people may be categorized as knowledge workers.

An other approach to the concept of information society, that has received interest within the field of information science, is dealing with information as a resource. An information resource is literally and concretely a source of information that has been established so that it can be reused (Levitan -82). As the traditional models are too heavily focused on information production (Järvelin & Repo -82), information as a resource approach stresses the vast amount of already generated information, which has been societally institutionalized for repeated use by one or many classes of users.

Information resources may be found in biological, physical, as well as in human context (Levitan -82). In the human context information resources take the forms of individuals, organizations, libraries and information centres, professional societies, office files, archives, documents, information systems and networks, television broadcasts, insurance companies, etc.

For the basis of our further information needs analysis we take a recent conceptualization by Horton (Taylor -82) of information system as a value-added process. This model has focussed on the contents of information systems and processes, and it covers the use of information, not only production and transfer.

The value-adding process contains four kinds of information which is needed for decision making: 1) data, 2) information, 3) informing knowledge and 4) productive knowledge. Raw data must be processed, i.e. value must be added before it becomes information. By selection, analysis, comparison and interpretation of information one gets informing knowledge. And finally, the process of moving from informing to productive knowledge is a judgement process, where options are presented and advantages and disadvantages weighed. At each step more value is added to the item under consideration.

This value-adding process may be evaluated from two different points of view, behind which one could name two separate interest fields in the information science of today:

- 1) information as a resource (IR), and
- 2) information as an attempt to deepen our understanding of our problems and the world¹⁾.

The first approach concentrates upon grouping, classifying and formatting data to form information into a baggage of information. An example of this is the life cycle of information production model which pays attention to information sources and resources becoming information products, services and systems. (Levitan -82).

Informing knowledge is already so much related to the use of information that it is seen among the IR people as a part of information marketing process. IR people admit the special features of information as a product. For example, Newman (Levitan -82) states that the information market is structured according, in addition to market-supply, to special infrastructural or institutional conditions containing technical, legal and political factors. Also Taylor's approach is IR-oriented

¹⁾ This approach having roots in phenomenology and hermeneutics has just started to gain interest within the information science.

although he pays attention to the information user, too, as a part of 'information environment' (Taylor -82).

At the organizational level the related phenomenon is information resources management (IRM). IRM is seen as a way of organizing information flows and information technology to accomplish organizational objectives by maximizing use, minimizing cost and assigning accountability (Horton -82). IRM attempts to manage and technically support the use of information within an organization, not so much from knowledge workers' point of view, but from an organizational point of view. However, this is already a big leap from information business much closer to the final users of information.

While the above described approach mainly deals with data and information, the other approach, information as an attempt to deepen our understanding of the world or our problems, deals with action and productive and informing knowledge. Most important here is the difference in attitudes towards the knowledge worker: while the IR people talk about information users in quite abstract terms, the 'understanding' approach starts its conceptualization from knowledge work action, task performing and the knowledge/information which is needed there.

While describing the value-added process model Taylor states, that in a decision process the final action is a personal one and the utilization of knowledge will depend not only on its validity, quality, or ease of access, but also on the degree of 'fit' between the knowledge provided and the environment within which the user operates and makes decisions (Taylor -82). So, Taylor notices the gap between markets of information products and the knowledge worker conducting his task. We think this gap should lead to deeper analysis of the knowledge work situations.

A typical knowledge work situation is a problem solving task, where we need information and its transformation to knowledge. A problem solving task may contain subtasks like interpretation, diagnosis, monitoring, prediction, planning and design.

which are examples of non-routine knowledge work and require expertise. Of course as parts of these tasks we have to acquire, process, store, transfer, etc. information.

A knowledge worker having a problem shapes first some kind of pre-understanding of the problem and its environment. If a problem is given from outside, someone perhaps gives him this first hand information. Now, all the background knowledge, beliefs, values, etc. from education, previous tasks and other experiences influence how he understands the problem. Advancement the problem solving case raises the need for problem - oriented information.

One may distinguish five kinds of knowledge and information in a problem solving task:

- 1) Knowledge that determines what the problems/tasks in the domain are (TK).
- 2) Domain knowledge, knowledge about the facts already known, informing knowledge (DK).
- 3) Problem knowledge, information about the conditions of the problem at hand (PK).
- 4) Problem solving knowledge, advising or productive knowledge (PSK).
- 5) Outcome knowledge, result knowledge (OK).

This conceptual approach to knowledge work tasks is reflected in real information needs and how the knowledge workers act while seeking information. It is important to "close" the process by outcome knowledge, because it will be shaping the other types of knowledge in the coming problem solving tasks, and it is worth noticing that this outcome knowledge is trans-

formed¹⁾ during the reporting process into result information so, that it can be added to the community information resources.

3. INFORMATION NEEDS AND SEEKING

As (Brittain -82) has pointed out, only a few recent user studies in information science have attempted to assess work situations and problems in detail and then to deduce from the findings the information needed to carry out work tasks. The cause is partially the background of the study area and the researchers and partially the conceptual frameworks structuring the study area. Frameworks like those by Paisley (-68), Allen (-69) and Line (-74) have stressed the importance of factors having influence on information needs and conceptualized information needs from the viewpoint of the users' (present) ways of seeking information. While important and true, these lists of influential factors have not provided us with conceptual tools to analyze the relation between information needs and the knowledge worker's tasks. More importantly, they give no tools for analysis of the effects of new information technologies and new services on their tasks and the associated information needs. Therefore, it should not be a surprise that these relations have been neglected in the information need studies.

Horton's recent paper (Horton -83) classifies information needs on the basis of Maslow's hierarchy of general human needs (e.g. biological, security, social, etc. needs). Horton also classifies information resources with respect to the hierarchy of information needs. While this model may be psychologically

1) This checking and cleaning while transforming may even destroy the knowledge from a document. As Streatfield (-83) has noticed during their EMIE-project certain official publications do not contain relevant information for their main users because of publishers decisions, what is suitable for publishing.

interesting and useful in structuring information resources of the general public, it does not give conceptual tools to analyze information needs from the perspective of knowledge work and evolving information technologies.

Our approach here is to consider information needs and seeking from the knowledge worker's problem solving perspective. Therefore we tie the information need and seeking concepts to the types of knowledge associated with problem solving. As a result we are able to present hypotheses on the effects of various types of technologies on information needs and seeking.

In performing their tasks, men are engaged in envolving practices. The evolution is philosophically, culturally, socially and technologically determined. In this paper we focus on the technological evolution. The practices are developed and maintained by professional communities through education, rules, standards, regulations, habits, peer evaluation etc. The practices reflect the state of knowledge and form the paradigm of the community under consideration.

Within the domain of each professional community, the practices prescribe

- the problems/tasks (TK),
- the working sequences and methods in solving the problems on performing the tasks (PSK),
- what information must be obtained about the object of study (PK),
- who is competent of performing the tasks (DK), and
- what the outcome looks like (OK).

These requirements are directly related to the types of knowledge in solving problems. We have to master these to a certain degree to become qualified members of the professional community. When determined by the community we call these the normative requirements on TK, PSK, PK, DK and OK.

Kunz and Rittel (-77) claim that the working patterns in all knowledge work and associated knowledge can be described to a great extent. This is supported by the evidence of expert systems design (Barr & Feigenbaum -82). Systems with performance comparable to that of human professionals have been designed (e.g. MYCIN, (Shortliffe -76), PUFF (Aikins & al. -82)). The methods and knowledge used in these systems resemble, and are equivalent to, those of human experts.¹⁾ Therefore we can, to a great extent, determine the normative requirements for any task/problem, although this may be a hard effort. Tools like EMYCIN (van Melle -81) for constructing expert systems seem to be a fruitful approach to ease the burden.

We want to stress that it is through the analysis of the practice, the working patterns, that we learn about the task requirements on various types of knowledge. The influential factors (e.g. Paisley -68) shape the practice and do not directly influence the normative requirements. According to Keen (Keen -80), however, there are no sufficiently good description methods available yet: the methods do not allow studies of the relations between the features of the task and the behavior of the knowledge worker, or the design criteria of knowledge work support systems. The level of description determines the possibilities of conscious redesign and evolution of, and support to, the task.

Within each discipline or mission there may be several different paradigms or practices in accomplishing the tasks. As they are contradictory, each of them may have unnecessary, inaccurate or even wrong requirements to be met. If these were distilled away, we would have the objective idea of the task consisting of the goals of the task and the means

¹⁾ It was learned in the GUIDON project that the inference procedure in the MYCIN type production systems cannot be easily learned by human students (Clancey -81). While equivalent to human inference methods they are not equal.

(i. e. steps and methods) to achieve them, and the objective requirements on TK, PSK, PK, DK, and OK associated with it¹). Our knowledge is limited but - as we believe - approaching the objective requirements. They can be found, never invented (Platonism). Being so defined, they are not meaningless to knowledge work support system design: they serve as the ideal for task orientation of the design work.

Returning back to the fallible human world and considering individual knowledge workers: following their community-influence practices when working on a problem/task they experience subjective information needs²). These are actualized step by step in the course of work depending on the approach and working method and their perception of the situation (current results, goals, possibilities, methods) at each step (Kunz & Rittel -77)³). Also these needs may be divided into

- the type and worth of a problem/task (TK),
- the problem solving methods (PSK),
- the facts about the problem (PK),
- facts about relations in the problem area (DK), and
- the requirements on the outcome (OK).

Some of these needs are met by memorizing the necessary pieces of knowledge. Others will be met by seeking the relevant knowledge from external sources. Some of the needs remain unsatisfied, especially if the knowledge worker can do without the needed knowledge. If not, the some of these needs are met

¹) We understand that the idea of each task belongs to the idea world, always subsisting as opposed to the time-dependent existence of material objects (Russel -59).

²) This is similar to information want' by Line (-74) and 'subjective Informations bedürfnisse' by Wersig (-73). However, we subdivide the concept by type of knowledge needed.

³) We do not consider his needs about the justification or ultimate effects of his task. So whether he is designing warheads or not is out of the scope of this study although it may influence his work (e.g. motivation and personal responsibility of existence of mankind).

by consciously creating the needed knowledge.

The workers' education and prior experience have a strong influence on the needs and the portions of them met by memorizing vs. creating vs. seeking and the portions not met at all.

Information seeking forms one part of knowledge work. It is the process of attempting to meet the subjective needs. Its role is supportive while the rest might be called the logistive (immediately furthering goal-achievement) part of knowledge work (Järvelin & Repo -82). There are also evolving practices in information seeking in all professional communities. These practices prescribe when and where to look for certain types of knowledge (TK, PSK, PK, DK, OK). In parallel to normative requirements on various types of knowledge these practices are norms on information seeking. They are reflected in part in the community's requirements on working environments for its members and in part in the services (e.g. handbooks, bibliographic retrieval systems) provided by the community to its members.

Each knowledge worker develops his individual information seeking practice by using his personal network of channels and sources of knowledge. These are organized by himself, his professional community, his organization and the information business. His practice forms the link between the type of problem/task and the type of knowledge needed, on one hand, and the channels and sources on the other. The practice differentiates between the channels and sources on the basis of subjective evaluation with respect to task and knowledge type, i.e. the worker's behavior in seeking problem solving knowledge differs from his behavior in seeking domain knowledge. His evaluation of the sources and channels covers issues like their capacity, speed, timeliness, coverage, selection quality, quality of transmitted information required efforts in using them, their flexibility, social factors (norms) etc. (Kunz & el. -77), (Christie -82). Of course this evaluation is rarely formal and explicit. Nevertheless, it is not only the principle of least effort that guides his behavior.

With this perspective, plain ranking lists of channel/source preferences should be augmented with information on types of knowledge sought, and types of problem or task under consideration. The influential factors listed by, e.g. Paisley (e.g. invisible college, reference group, cultural system) are important norms in forming the individual information seeking practices.

The way of using individual knowledge sources or channels and the questions posed on them (demand (Line -74)) depend on the structural and functional requirements of the source/channel as well as on the user's mental model of the channel/source. The responses by the channels/sources influence the user's evaluations of them and, indirectly, the norms on information seeking of the user's professional community.

4. ON THE EFFECTS OF INFORMATION TECHNOLOGY

The purpose of information technology is to augment and support knowledge work. By knowledge work augmentation (Engelbart -62, Engelbart -75) or support we mean increasing the knowledge worker's understanding of, and capability to solve, complex problem situations. This results in speedier, better or novel solutions to, and understanding of, his problems.

There is a mess of all kinds of tools for knowledge work. The tools are based on various technologies (e.g. fiber optics, VLSI (very large scale integrated circuits)), they support various functions (e.g. observation, symbol manipulation), they have varying domain-orientation (e.g. specific tools for medical diagnosis vs. general text-processing tools), they operate on varying degrees of expert knowledge (e.g. simple data retrieval vs. expert systems), they have a variable user-orientation (e.g. customized vs. general tools), and they are meant for varying numbers of users (e.g. single-user tools vs. tools for cooperative team work).

The tools also have diverse effects on knowledge work: some automate formerly manual phases of work, some substitute new methods for former ones, some allow optimization where formerly only "muddling-through" was possible, some give hard-and-fast solutions to former critical problems and so allow the evolution of problems, some just gather and organize formerly available data or information so that access and processing becomes possible or easier, some just make other tools obsolete (Järvelin & Repo -82).

In general, the tools increase knowledge worker's effectiveness by

- increasing his personal efficiency
- speeding up problem solving
- supporting communication among workers
- supporting learning and training
- increasing the possibilities of planning and
- controlling work. (Alter -80).

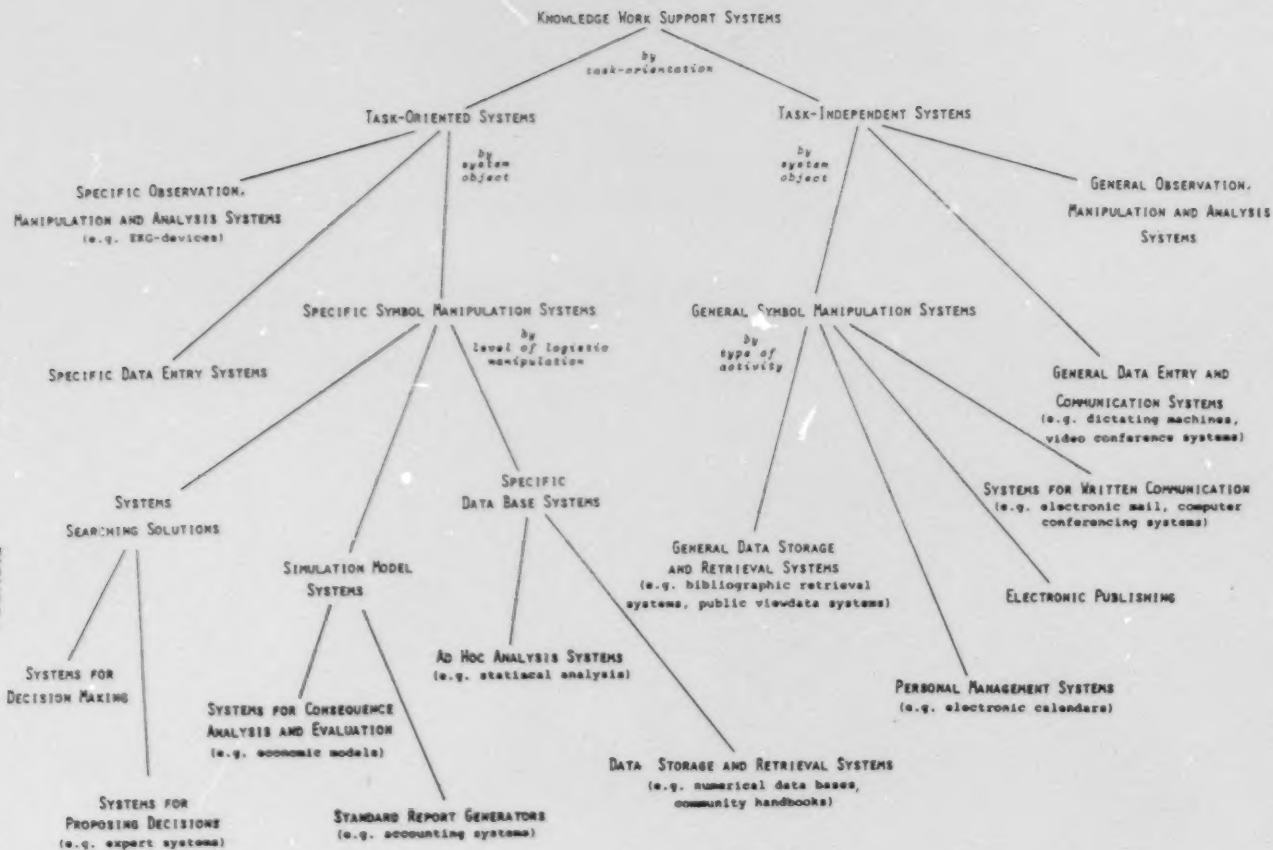
To describe and to analyze the effects of the tools on information needs and seeking we outline below a taxonomy of the tools and discuss then the specific effects in three cases: computer conferencing systems, data base systems and expert systems.

4.1 A taxonomy of tools for supporting knowledge work

We approach the development of the taxonomy (Fig. 1) from the viewpoint of an individual knowledge work task. We can then firstly classify some tools as logistic, or task-oriented, providing support in the logistic parts of the task (e.g. finding out a patient's disease, computing the stress on various parts of a bridge under design). The rest of the tools are task-independent, providing support in the management and administration of the task (e.g. organizing documents, time-accounting, communication). We must tie this classification to the task, because some tools are logistic in one case and task-independent in the other. Consider, for example, an IR-system from the perspectives of an information officer vs. a chemist.

8/45

FIGURE 1.



The next level contains a classification of the tools of either type into symbol manipulation systems, data entry systems and observation, manipulation and analysis systems. This classification is based on the system object: symbols, symbol production, and physical objects and phenomena. Systems of the last type support our senses and the manipulation of the object of work. Information technology deals mainly with the first two classes. Examples of data entry systems are various specific sensor systems, dictating machines etc. We focus on the first class. Real systems may, of course, cross these class boundaries (e.g. a process control computer).

The classification of the logistic symbol manipulation system has been built on Alter's taxonomy of decision support systems (Alter -80). It is based on the degree of tool intelligence in the task. Specific data storage and retrieval (DSR) systems support storing and retrieving task specific data and knowledge. Systems that make decisions for the knowledge worker know what decisions are to be made, what are the relevant data and how they are processed. So the latter type systems have more knowledge about the task, i.e. are more intelligent than the former type systems.

Within the general symbol manipulation systems we distinguish general data base systems, systems for personal management, electronic publishing and systems for symbolic communication. This is based on the respective activities in knowledge work.

We believe this taxonomy creates a partial order within the mess of all kinds of tools. We think the taxonomy is fruitful for analyzing the effects of information technology. It may be difficult to place some tools into the taxonomy. This is so partly because real tools may cross the class borders and partly because we have simplified the real multidimensionality of tools and the support they provide.

Consider, for example, the concept of task-orientation of a tool. From the logical point-of-view a tool may be task-oriented because

- the data and information are selected, gathered and organized strictly for the task; domain knowledge is generated (this restricts the domain of application of the tool),
- the tool incorporates methods (e.g. statistical analysis) required in the task (this is methodological task-orientation),
- the tool's conceptualization of the object is suitable for the task (this is conceptual task orientation)

To illustrate the differences of the first two, consider the EMYCIN tool for building expert systems. It is domain-independent in a way, because it may be used in building expert systems for various fields in medicine, and in structural engineering. However, it is methodologically limited so that the problems for the experts must be deductive (analysis and consultation) problems. EMYCIN is not suited for building experts for formation, simulation or planning problems. (Van Melle -81).

The task-orientation of a tool may also be due to its responsiveness to the physical or representational requirements of the task (e.g. speed, shape weight; media: audio, video; verbal, tables, diagrams). The taxonomy does not show, either, the dimensions of technology, user-orientation nor cooperativeness.

It should be obvious that all classes of the taxonomy are not relevant to all tasks, and that there are relevant classes without existing tools for some tasks. The classes reflect the present tools and their connections.

4.2 The study of three supporting systems

In practice there are several knowledge work support systems for use in our working environment. Our aim is to characterize the role of support systems for information needs and seeking. We have picked up three support areas gaining vast interest among those who develop these systems:

- 1) Computer conference systems,
- 2) Database systems, and
- 3) Expert systems.

Computer conference (CC) is asynchronous communication with typed messages from an online terminal or work station; examples of already internationally used systems are the Electronic Information Exchange System (EIES) in New Jersey and COM in Stockholm. As tools in receiving information CC systems locate somewhere between an electronic mail and an information retrieval system, and they support especially group communication (Hiltz & Turoff -78), (Palme -81), (Repo -83).

CC-systems do not change, at least directly, the logistic working practice and so plain changes in subjective needs and normative requirements of knowledge should not be expected. These systems are tools for filed communication. But they are not indifferent towards the types of information they provide.

To take an example, what is the difference between reading the conference proceedings and attending a conference? In a conference you gain knowledge about the new hot problems, in the area (TK); what are the relevant approaches & methods (PSK); and the newest domain facts (DK). In comparison to reading the articles, TK and PSK are much better available in the discussions when attending a conference. We argue that this is even accelerated through asynchronous CC-systems, though face-to-face meetings still have their role, e.g. in solving disagreements. - So, it should be investigated whether the use of CC-systems direct PSK-needs to the CC-systems participants and whether prior PSK-sources are less utilized (those colleagues, who are not in the systems used) after particular CC systems are in full use within a certain group of professionals.

To reiterate the hypothesis, CC systems affect the personal information seeking practices directly so that attempts are made to obtain especially TK and PSK more through the CC-systems and less through the former channels/sources, if the use of the systems is convenient. In addition, availability of CC systems gradually changes the community norms on information seeking and working environments, and indirectly they may accelerate the evolution of the logistic practice. The last hypothesis should be studied more thoroughly from group communication practices' point of view.

Next we discuss the potential effects of data base systems. We shall concentrate on data storage and retrieval (DSR) systems¹⁾. Some DSR systems contain bibliographic data, some factual information (DK)²⁾ about some domain (DK-DSR systems, e.g. numeric data bases), some specific observation and measurement data/information about the object at some knowledge work process (PK-DSR systems), and finally some contain problem solving knowledge in the form of how-to-do advice (PSK-DSR systems, e.g. advice in a community handbook (Engelbart -75).

The bibliographic DSR systems are likely to cause minor changes in information seeking practices. This is due to publication delays and the fact that literature transfers mainly OK. The PK-DSR systems are rather likely to reflect changes done elsewhere in the working practice than to be agents of change. We concentrate below on DK-DSR and PSK-DSR systems. All DSR systems affect individual information seeking practices because they gather, organize, and provide access to data/information-/knowledge. However, there are obvious differences with respect to the type of knowledge.

¹⁾ In fact these are not just data storage & retrieval systems but data/information/knowledge storage and retrieval systems. But for shortness: DSR systems.

²⁾ Actually, information about facts, which becomes DK when used in problem solving.

The DK-DSR systems quite obviously affect the DK seeking practice. Such a system may replace a set of handbooks, data tables, individual archives, many phone calls and piles of journal articles. The effect may be strong if the system is easy to access and if the data/information, their measurement procedure and representation match the task requirements. This need not be so because the data are always theoretically contaminated and therefore possibly of no value in the task.

The DK-DSR systems do not directly change the logistic working practice and therefore neither the normative requirements nor subjective needs.

The PSK-DSR systems may be used to find advice about how to go on with the task/problem when uncertainty arises, or just to confirm our ideas about the next steps. Such a system does not do anything for you, it merely allows you to ask for advice. It certainly changes the user's PSK seeking practice: on problems in its domain the user is likely to consult the system and use less his old PSK channels and sources. Through its advice the system also shapes the user's logistic practice and his subjective needs on PK, DK (and perhaps even PSK not covered by the system) and PK, DK seeking practice. The system itself may be characterized as a storage of normative requirements.

Consider finally the effects of an expert system that incorporates DK, PSK, knowledge about what kind of PK is relevant and some explanation and tutoring capability. Generally speaking, an expert system brings expertise to those not having it yet. This means rationalizing the working patterns to the level of human experts in the domain. This results in (partial) reorganization of the whole task, i.e. developing a new practice both for the logistic and information seeking parts of the task. New methods, new knowledge sources new channels need to be organized. This is set forth at the installation of the expert system through training.

It is obvious that new normative information requirements and new subjective needs develop - those of the system developer and the system user community have their effects. While the new

user was not an expert previously, he does not need to be an expert later, either. However, the user-system symbiosis may approach the objective requirements of the task better. But the user must be educated enough to consider when the system, or which system if there are several, should be used (TK) and to understand the results and its limitations (OK). The used expert system should also be updated continually or it becomes a closed system losing its significance.

The most marked changes are at the subjective needs of PSK and individual PSK seeking practice: the user knows what to do and stops using old PSK channels and sources.

5 DISCUSSION

In this study we discussed the strain we can discover between "information as a resource" and "information as an understanding tool". Our approach was based on the last mentioned, dealing with information and knowledge requirements of knowledge workers when conducting their tasks. We considered knowledge work as a problem solving task creating objective, normative and subjective knowledge requirements and those reflected in information seeking.

From our information need and seeking view we generated a taxonomy for knowledge work support systems. These systems are constituted of technology, information and knowledge and human intelligence. While pure technology and general information may only support some information seeking routines, knowledge bases of expert systems and human communication augmentation tools reach a much higher level of support. It seems that practical support systems contain intelligence tied to both information technology and human colleagues. This arrangement makes the system flexible and open to the changing knowledge work environment.

The framework and the three cases we have discussed suggest the following general hypotheses

- all support tools change the knowledge worker's information seeking practices, because they influence the environment directly where the work is done,
- task-independent tools do not directly affect the logistic tasks and so they have no direct effects on the subjective needs of knowledge: only limited success in increasing the knowledge worker's effectiveness can be expected from efforts to develop task-independent tools, or at least it is hard to point out
- evolving individual working and information seeking practices, and evolving subjective information needs gradually change the community's practices and norms.
- task-orientation is created by task-oriented information and knowledge normative surroundings. e.g. organization and logistic symbol manipulation capabilities.
- much of the outcome knowledge remains in the knowledge workers' heads and in the task oriented tools: it is not distributed through literature or other IR-market.
- all tools have indirect effects on knowledge work, information needs and seeking, which are difficult to track down.

The situation challenges the information science communities to gain broader understanding of knowledge work support systems: two decades' works of artificial intelligence dealing with expert systems are coming out from research laboratories, and on the other hand, new tools for communication have already arisen attention within the field of information science, e.g. (Hiltz & Kerr -82). By understanding the nature of knowledge work and its information requirements, information scientists

could be building bridges between general information systems and services, and knowledge required while conducting knowledge work tasks; it is worth noticing that we cannot build a bridge, if we only know the other edge of the river.

For the practioners of information services this development promises changing professional requirements. Information specialists have to broaden their expertise from general online databases to knowledge bases and communication systems, instead of only intermediating information they have to be able to process it with the user, they have to take part in planning and installation of knowledge work support systems in organizations, etc. Also librarians, archivists, publishers etc. have to consider the differing information needs more exactly, and this will lead to, e.g. on demand publishing and stressing the content of information instead of covers, documents. There have been discussions on the service of the information end-user already a long time; we information scientists are just starting to realize what this really means. And when we shift from theory to practice, we will learn how impossible our task is. This recovery should teach us to construct such knowledge work support systems that are technically and economically possible and socially and politically desirable.

REFERENCES

- T. R. Addis, Expert systems: an evolution in information retrieval, *Information Technology: Research and Development* (1982), pp. 301 - 324.
- J. S. Aikins & al., PUFF: An expert system for interpretation of pulmonary function data. Report No. STAN-CS-82-931 (Depts. of Medicine and Computer Science, Stanford University, Stanford, CA., 1982).
- T. J. Allen, Information Needs and Uses. In: C. A. Cuadra (ed.), *Annual Review of Information Science and Technology*, vol. 4(1969), pp. 3 - 239.
- Steven L. Alter, *Decision support systems: Current Practice and Continuing Challenges* (Addison-Wesley, Reading, MA, 1980).
- A. Barr and E. A. Feigenbaum, eds., *The handbook of artificial intelligence*. Vol. II (Kaufmann, Los Altos, CA., 1982).
- J. M. Brittain, Pitfalls of User Oriented Research, and Some Neglected Areas, in I. Friberg, ed., *Proc. IRFIS 4*, 14.-16. Sept. 1981, Borås, Sweden (Högskolan i Borås, Borås, 1982), pp. 213 - 227.
- B. Christie, Face to file communication. A physiological approach to information systems (John Wiley & Sons, New York, 1981).
- W. J. Clancey, Methodology for building an intelligent tutoring system. Report No. STAN-CS-81-894 (Dept. of Computer Science, Stanford Univ., Stanford, CA., 1981).
- S. Crawford, Information Needs and Uses, in: *Annual Review of Information Science and Technology* vol. 13 (1978) (ASIS, Washington D.C., 1978), pp. 61 - 81.
- R. Davis, Interactive Transfer of Expertise: Acquisition of New Inference Rules, *Artificial Intelligence* 12 (2) (1979), pp. 121 - 158.
- P. F. Drucker, *The Effective Executive* (Harper and Row, New York, 1966).
- D. C. Engelbart, Augmenting Human Intellect: A conceptual Framework, Summary report, project no. 3587 (Stanford Research Institute, 1962).
- D. C. Engelbart, Coordinated Information Services for a Discipline or Mission-oriented Community, in: R. L. Grimsdale, F. F. Kuo, eds., *Computer Communication Networks* (Noordhoff, Leyden, 1975).

R. P. Futrelle, L. C. Smith, Expert systems using semantic representations of full text for question answering. Proceedings of the ASIS Annual Meeting, Vol. 19(1981), pp. 103 - 106.

S. R. Hiltz, E. B. Kerr, Computer-mediated communication systems, status and evaluation. A volume human communication research series. (Academic Press, New York, 1982). 180 s.

S. R. Hiltz & M. Turoff, The network nation. Human communication via computer (Addison-Wesley Publishing Company, Inc., Ontario, 1978).

F. Horton Jr., IRM: the invisible revolution. Computer World (16)(1982), 7 p.

F. W. Horton, Information literacy vs. computer literacy, Bulletin of the American Society for Information Science 9(4) (1983) pp. 14 - 16.

K. Järvelin, A. J. Repo, Knowledge work augmentation and human information seeking, Journal of Information Science 5(6)(1982), pp. 79 - 86.

P. W. G. Keen, Adaptive design for decision support systems, Data Base 12(1-2) (1980) pp. 15 - 25.

W. Kunz, H. Rittel, Zur Logik von Forschung und Dokumentation, Die Naturwissenschaften 55 (8) (1968), pp. 358 - 361.

W. Kunz, H. Rittel, A Systems Analysis of the Logic of Research and Information Processes (Verlag Dokumentation, München, 1977).

W. Kunz, H. Rittel, W. Schwuchow, Methods of Analysis and Evaluation of Information Needs: a Critical Review (Verlag Dokumentation, München, 1977).

K. B. Levitan, Information Resources as "Goods" in the Life Cycle of Information Production, Journal of the American Society for Information Science 33(1)(1982)1, pp. 44 - 54.

M. B. Line, Draft Definitions, Aslib Proceedings 26 (2) (1974), p. 87.

W. J. van Melle, System aids in constructing consultation programs (UMI Research Press, Ann Arbor, Mich., 1981).

H. Menzel, Information Needs and Uses in Science and Technology, in: C. A. Cuadra, ed. Annual Review of Information Science and Technology, vol. 1 (1966) (ASIS, Washington D. C., 1966), pp. 41 - 69.

W. Paisley, Information Needs and Uses, in: C. A. Cuadra, ed., Annual Review of Information Science and Technology, vol. 3(1968) (ASIS, Washington D.C., 1968), pp. 1 - 30.

J. Palme, Experience with the use of the COM Computerized Conferencing system (Swedish National Defence Research Institute, Stockholm, 1981).

B. Russel, The problems of philosophy (Oxford U.P., Oxford, 1959).

A. J. Repo, Computer conference in information service, Research reports, Technical Research Centre of Finland, (191)(1983), 50 p.

E. H. Shortliffe, Computer-based Medical Consultations: Mycin (Elsevier, New York, 1976).

V. Stibic, Personal Documentation for Professionals. Means and Methods. (North-Holland Publishing company, Amsterdam, 1980).

V. Stibic, Tools of the mind. Techniques and Methods for intellectual work (North-Holland Publishing Company, Amsterdam, 1982).

D. Streatfield, Moving towards the information users some research and its implications. Draft to Social Science Information Studies (1983) 39 p.

R. S. Taylor, Value-Added Processes in the Information Life Cycle, Journal of the American Society for Information Science 33 (5) (1982), pp. 341 - 346.

G. Wersig, Informationssoziologie: Hinweise zu einem Informationswissenschaftlichen Teilbereich (Athenäum Fischer, Frankfurt (aM), 1973).

INFORMATION SCIENCE RESEARCH

VERSUS THE PRACTITIONER

A.R. Blick

INFORMATION SCIENCE RESEARCH

VERSUS THE PRACTITIONER

A.R. Blick

Abstract

In this opinion paper the relationship between the information science researcher and the practitioner is discussed. Reasons why the practitioner does not use research results are given and areas for further research in the author's own information service are proposed. The conclusions drawn are a) there is an undesirable reaction against research by the practitioner, b) there is a serious lack of funding of the development phase, c) there is insufficient mission-oriented research.

Quotations

When I knew I was going to give this paper I took the opportunity of the recent IIS Conference to ask the managers of several large industrial information departments to give me an instinctive reply to the question "what do you think of information science research?" Their replies are given below.

"Information science research is irrelevant to my work as a practitioner".

"Computer technology advances are rendering information science research obsolete before it has been completed".

"Industrial research is nowadays highly mission-oriented. Why is this not the case with information science research?"

"Someone should write a paper on how to translate papers on information science research into an understandable form".

"Information practitioners should involve themselves in information science research in order to influence it and exploit it".

"Don't talk to me about information science research!"

"Information science research is very variable in its value but I wonder if some of it ever will be relevant".

"Some industrial information departments in the United Kingdom have budgets larger than the total amount spent on information science research in the U.K. Is it fair to expect much from research from such small resources".

Introduction

I obtained these quotations before I was aware of the recent Aslib publication titled 'Research and the Practitioner'.⁴ However, the hostile nature of the quotations is reinforced in the Aslib publication by many other quotations of a similar nature and the authors say "At times it would almost seem that there is an instinctive resentment of research amongst practitioners". The paper I will present is an opinion paper and is not based on the results of such a detailed study. However, it is deliberately titled 'Information science research versus the practitioner' because that is how the situation

appears to me after many years in information work. Instead of there being a unified or sympathetic approach between the information science researcher and the practitioner, there seems to be an antagonism based on the practitioner's belief that what the research worker is doing has no relation whatsoever to what the practitioner is doing. Regardless of whether or not this is justified, it is a widely-held belief. My own first substantial contact with information science research occurred when I attended the IRFIS meeting in Oslo. This was the first information science research meeting I have ever attended and it came as a cultural shock. I was appalled that I knew so little about the research work that was going on and I was also appalled that I could barely understand half the words that were used. I realised that none of the work described had any immediate relevance to my own activities and I also learned that the research worker in information science believes that practitioners never read research papers and are ignorant of the value of the research work being done. Certainly the two groups are worlds apart and this is indicated by the fact that I was the only practitioner at IRFIS 3 and am probably the only one here today.

This is obviously undesirable from many points of view, not least being that £1.5-2 million is spent in the U.K. alone on information science research. Such an investment should generate something of direct value, other than knowledge for knowledge's sake and it is because so little direct value seems to accrue that I am here today talking on this topic.

Before going further into the paper I would like to point

out that although I am a practitioner in an industrial information service I have been involved in what might be loosely called 'research' into the operations of the service.

For example, within the Beecham Pharmaceuticals Research Division information service we have conducted two substantial investigations of our information bulletin. The first¹ concerned the value of the bulletin itself and the second² compared the efficiency of the bulletin with the alternative of on-line SDI services. We have also constructed a value ranking list of our journals and have done journal use studies (both unpublished) and we have looked at the effect of the introduction of on-line search facilities on the choice of search tools used by information scientists.³

In all these examples, the purpose was to obtain information in order to make decisions about the service. For instance, the first study of the information bulletin investigated the coverage of relevant items in the literature and the importance of immediacy of notification of these items. It was discovered that 9% of items in the bulletin were considered vital by customers and 36% important but 94% of these vital items would not reach the user in time without the bulletin and 59% would be missed altogether. Of the important items, only 34% would be seen in time and about 59% would be missed altogether. A cost benefit for the bulletin was demonstrated of 4.6 x the cost of the bulletin - one of the first practical cost benefits demonstrated I think.

The second bulletin study showed that the information bulletin retrieved only 61% of the total relevant references

which were found in computer SDIs run on on-line databases for 6-months. It also showed that the on-line SDIs found only 43% of the relevant references in the bulletin for the same period.

Both these studies were highly mission-oriented and demonstrated it was correct to have such a heavy investment of effort by our Information Services in the information bulletin.

So, although I am speaking at this meeting as a practitioner I can claim a little understanding of the complexities of information science research. My contact with the research activity has made me sympathetic to the research workers' activities and has heightened my concern at the gulf between practitioner and researcher.

Why Practitioners Do Not Use Research

I will now enumerate what I think are some of the reasons why practitioners do not make use of research findings. I must stress that these are personal opinions drawn from discussions with information practitioners in the industrial sector. Table 35, page 57 of the Aslib publication is also of interest on this topic but does not specifically consider why practitioners do not use research results. It also relates predominantly to librarians, (82% of respondents were members of the Library Association).

1. Poor Communication - I think it is inarguable that there is poor communication between research workers and practitioners. I know considerable thought has been given to ways of improving this communication and of improving

the dissemination of research results and this has been achieved to some extent. However, practitioners do not attend information science researchers' meetings, and vice-versa, and I suspect they do not read each others' journals. They are two distinct communities within the information world. This is a ridiculous state of affairs as the total information community is quite small.

2. Literature Overload - Even if research results are disseminated more widely I doubt if this will improve the situation very much. Already the practitioner is inundated with library and information science journals and also with journals directly related to his primary activities such as those concerning on-line searching. My own unit contains some 20 feet of manuals that an information practitioner must read and understand in order to do on-line searching. The end result is that the practitioner has to concentrate on the literature which contains hard, practical information of direct value to him and ignore the journals which carry research results.
3. Jargon - The communication of research results is hampered by the jargon used by the research workers. At least this is how it is seen by the practitioners who are so out of touch with research work that they cannot understand many of the terms used. As a result it becomes very hard work to read publications such as the Journal of Documentation and this is an added reason why they concentrate on the

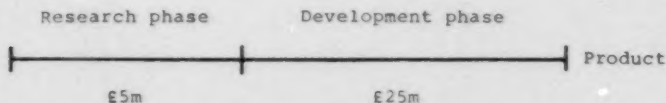
practical journals.

4. Technology - Information technology has never progressed so quickly as it is doing now. The practitioner is kept at full stretch striving to keep up-to-date with the new technological developments and with their practical application. Examples are such things as on-line literature searching, on-line databanks, video discs, teletext, computer software packages, computer hardware, (the move to microcomputers is causing a considerable re-think for practitioners), microform hardware and software, chemical structure modelling, electronic journals, etc. The practitioner is under such pressure from these developments that there is little time to look at research results. In addition it seems in many cases that it is the new technology which brings the new information systems rather than information science research itself generating them.
5. The Service - The priority of the practitioner is to provide a service to the customer. Information departments are rarely staffed to the level the manager would like and so are usually over-stretched. As a result there is little or no time for the practitioner to do research and time is rightly concentrated on meeting customers' requirements. There is equally little time for the practitioner to pick up a research result and convert it into a practical application.

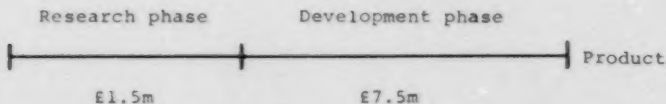
6. Resistance - There is undoubtedly a natural resistance to change intrinsic in mankind and this goes for information practitioners as well. Consequently there is rarely an eagerness to pick up new research ideas and there is considerable temptation to shut one's mind to them. There is some justification in this because, generally speaking, it takes considerable effort to change what may be a long established system onto a new basis shown to be better by research results. For example, if an in-house document records store is based on a UDC system and research shows that a thesaurus-controlled textual approach would be better, the amount of time and effort to change would be considerable. In an overstretched unit, this is unlikely to occur and the practitioner in charge will ignore the research finding.
7. Education - The educational background of many of the present information practitioners in management positions in industrial information units includes very little contact with information science research. Not having had early contact with this research they are, therefore, not all that sympathetic to it. I firmly believe that this situation is improving as the academic education of information scientists frequently includes research education nowadays. I support wholeheartedly any move to make our information science students more research aware.
8. Development - All these reasons are valid in themselves

and contribute to the problem. However, I think the main problem lies in the very large gap between a research worker obtaining research results and these results being translated into a practical application. I would like to draw the comparison with the development of a new drug. This takes place in three phases. First, the fundamental research which leads to the discovery of the promising active compound. Second, the development phase which provides all the evidence that this drug from the research phase has the properties and safety required, and third, the marketing of the drug. Of these three activities, the development phase is substantially the longest and most expensive. If it takes two years for the research phase, it can take up to eight years for the development phase. If this is translated into information science research results it is immediately apparent that, as there seems to be no-one doing development work, the gap between the research result and the application of that result is virtually an insurmountable hurdle.

New Drug



New Information System



A typical new pharmaceutical product would cost of the order of £5 million in the research phase and £25 million in the development phase. If this ratio of development cost to research cost is transferred into the information science research area then the £1.5 million being supplied by the British Library Research Department in the U.K. should be supported by £7.5 million for the development of that research work into usable products. The development phase of drug research is inordinately expensive and the comparison made is unrealistic but it does emphasise the point that substantial funds and effort are required before the present information science work being undertaken can be put to use.

The problem is not, of course, unique to information science research. Only this year the Advisory Council for Applied Research and Development (ACARD) in the United Kingdom reported that £30 million should be spent to overcome the barriers between academic research and the exploitation of this research by industry. In the information area, at the recent Institute of Information Scientists Conference at Oxford, the Minister for Industry and Information Technology said £5 million would be available for providing up to 75% of the cost of new information systems such as new databases.

The problem therefore is not unique to information science and nor is it being totally ignored but it still remains

the major problem.

Whether or not the research worker is the person to undertake this development work is another question.

Where This Practitioner Would Like Research

In order to develop this part of my talk I will have to give some background on the services provided by my own information services department.

Information services departments in some industrial organisations are very large and constitute a substantial overhead on that organisation. Information department budgets of £1.5 million per annum are not unknown. The budget for Information Services within the Beecham Pharmaceuticals Research Division is about one third of this figure but is still appreciable. The majority of the expenditure is used in the traditional information activities of current awareness notification, retrospective searching, library purchases of books and periodicals and in-house document systems. The Research Division consists of 5 separate sites, all of which have information units which offer exactly the same services to the users. It is my role to co-ordinate all these activities and to ensure that the Research Division has an information service which meets its requirements. Its requirements are dictated by the needs of the scientists who are performing research into new therapeutic products. In consequence the information scientists employed are all science graduates, several of whom have information science qualifications also.

In total, there are 10 professional, and 14 clerical, staff working in the information service. Each site information service has a unit head and these unit heads, together with an information systems specialist, meet periodically to review and progress the total information service. It is at such meetings that information science research findings should be discussed.

I readily admit that the information service provided to the Research Division is not perfect. This is due to many factors, one of which is lack of knowledge of some basic parameters essential to decision-making in an information service. I would like to see research done in order to provide this knowledge. The research would be mission-oriented, aimed at specific problem areas and parameter definition. I will give 8 illustrations of the areas. Research may already be being performed in these areas but if so, has not come up with definitive answers.

I would also point out that the Aslib publication gives a ranked list of 38 areas for future research effort but that the areas were pre-ordained in the questionnaire. My list is much shorter and, of course, is highly personalised.

1. Information Communication - One piece of research work I was very impressed by was that of Allen of MIT⁵ when he looked at the information transfer networks in the organisation and he came up with the idea of information gatekeepers. There are other research results around which suggest that libraries play a small role in the information transfer process within an organisation. I have an

intuitive feeling that Beecham Pharmaceuticals Research Division Information Services play a substantial role in this process largely because of the active current awareness services. I would be pleased to see research results on the information transfer process in an organisation where a very active information service is provided.

I would also like to see such results compared with an information service which concentrates on developing and providing facilities for the users to use themselves. In other words, to compare the information transmission efficiency of a service which services the users extensively against a service which facilitates do-it-yourself. There is considerable talk in the information profession of a move towards users coping with their information needs on a do-it-yourself basis as computer technology advances make this possible. It may or may not be beneficial for the user but before the users demand it should happen I would like to know the advantages and disadvantages in terms of efficient information retrieval and transmission.

2. Current Awareness - Following on from this first area I would like to know the real value of current awareness services and how varying these services affects their value. We put a considerable amount of effort into current awareness and all our performance measurements

suggest it is valuable and very worthwhile. Traditionally, current awareness has also played a major role in the activities of the information profession. Do we really know its true importance and do we know the most effective way of disseminating current information?

3. What Services - From the findings in (1) I would like to know where Information Services should be concentrating its activities. Information departments are traditionally based on current awareness and retrospective literature searching. Is this the best utilisation of information scientists or should they be performing some totally different activity to obtain maximum exploitation of information?
4. In-House Information - From conversations with managers of other industrial information departments it is apparent that in-house generated information which is then stored in a central collection is utilised very little by customers. It is very surprising that in-house information which must, by its very nature, be highly relevant to research workers, is so little used by them. Is it that the nature of the user's work is such that he is always looking forward and so needs to refer back very rarely? Does he value externally-published work more highly than the work of his own colleagues? I would be interested to know why in-house document collections are so little used and whether changing their storage format would change the

frequency of use.

5. Do-It-Yourself - Current opinion in industrial information services is that there will be a progression towards do-it-yourself for customers. This has been mainly brought about by the onset of interactive computer facilities. At the present time, with the present services and technology, there are the strong practical reasons why do-it-yourself is not appropriate. However, this will change in the future but is it desirable in itself that the intermediary role of the information specialist should remain or should disappear? What will be the future role of the information specialist? This is obviously a very important question for the future.
6. Categories of Customers - I spoke earlier about different phases in the development of a new drug. There would appear to be some evidence that the information requirements of users working in the research phase are different from those in the development phase. It would be useful to know how and why the information requirements of a research worker doing fundamental research differs from one doing developmental research and again differs from a user in the management area. Which information needs are greatest for each of these categories of user and how should they be met?
7. Performance Measurement - Performance measurement is

highly advisable for any service activity. It is something that most industrial information services can do to some extent. There is some literature around on performance measurements but it would be very useful if this could be strengthened and guidelines produced on this topic. What the industrial information practitioner is looking for is information on what he should be measuring, how it can be measured, what decisions to make on the basis of the results, and how all this can be achieved with the minimum of fuss and effort.

8. Reading Habits - Within our environment which contains a comprehensive current awareness service the majority of customers expressed a preference for non-circulation of journals. This ensured journals were always available in the library. The impact of this decision on the transmission of information could be very substantial and may or may not be favourable. This has lead me to feel the need for hard information on how users obtain information from the current literature and what factors influence it.

Conclusion

In this opinion paper I have tried to show why information practitioners do not use information science research results but at the same time to show that research work is necessary to the practitioner. There are obviously a variety of reasons why there is this gap between the researcher and the practitioner. However, the gap would appear to be caused by

two main reasons:

1. There is little or no funding, and consequently little or no work, in the development phase of information science research. At present the researcher cannot progress his work into a product and the practitioner is understandably unwilling to take-up ideas still in the research phase;
2. Little research is being done which can be considered mission-oriented. Many questions exist in the practical world of information science which are not being answered by the research worker.

I would like to suggest that the information science research profession should do two things. First, it should campaign amongst its financing bodies for funds to support development work and second, it should concentrate some of its research in areas which would be directly helpful to the information practitioner. Success in these two activities could then lead to a close and valuable relationship between the information science researcher and the information practitioner.

References

1. BLICK, A.R. et al *The Value of a Weekly In-House Current Awareness Bulletin Serving Pharmaceutical Research Scientists* *The Information Scientist*, March 1975 19-28
2. BLICK, A.R. et al *A Comparison of On-Line Databases With a Large In-House Information Bulletin in the Provision of Current Awareness* *Journal of Information Science*, 4(2/3) May 1982 79-86
3. BLICK, A.R. et al *The Effect of the Introduction of On-Line Facilities on the Choice of Search Tools* *The Information Scientist*, 12(1) March 1978 25-31
4. LYNAM, P. et al *Research and the Practitioner* Aslib Occasional Publication No. 27, 1982
5. ALLEN, T.J. *Organisational Aspects of Information Flow in Technology* Aslib Proceedings, 20(11) November 1968 433-454

METHODOLOGICAL PROBLEMS IN USER STUDIES, AND THE
ORGANISATION OF IN-HOUSE LIBRARY RESEARCH

by

F.C.A. Exon, Research & Development Librarian/Lecturer
Western Australian Institute of Technology

METHODOLOGICAL PROBLEMS IN USER STUDIES, AND THE
ORGANISATION OF IN-HOUSE LIBRARY RESEARCH

by

F.C.A. Exon

Research & Development Librarian/Lecturer
Western Australian Institute of Technology

Abstract

This paper contends that User Studies has developed a profound methodological crisis in the apparent opposition of quantitative and qualitative methods. Much of this argument is directed towards only academic researchers. However, its increasing presence in the professional literature of librarianship and information science has forced in-house library researchers to consider the basis for choosing between these methodological schools.

This paper also argues that in-house library research is the interface between academic research and professional practice, and that while methodological problems cannot be ignored, the cost-effectiveness of data collection techniques is of primary concern. As a contribution to this aspect of the problem, an attempt is described to ascertain the unit costs of data collection in two new studies recently conducted at W.A.I.T., one using classic social survey quantitative methods, and the other using those of "Illuminative Evaluation".

METHODOLOGICAL PROBLEMS IN USER STUDIES, AND THE
ORGANISATION OF IN-HOUSE LIBRARY RESEARCH

by

F.C.A. Exon, Research & Development Librarian/Lecturer
Western Australian Institute of Technology

Introduction

There is inevitably a gulf between theory and practice in information work. The daily pressures of the real world compromise principle in the pragmatic interests of service and survival, while theoretical work often proceeds best when liberated from the distractions of the operational environment. This creative separation of functions has distanced the researcher from the practitioner. Perhaps this is acceptable when times are good, but it is dangerous in a period of financial stringency. Poverty causes envy and protectionism. Yet research ("systematic investigation made public") is essential to the flexible management demanded by our present circumstances.

Consciousness of research might be regarded as the hallmark of professional maturity; certainly there is a great demand for research from the practitioners. For example, a participant in a recent American Seminar on research observed :

"Too often the library profession takes what could be called a lumber yard approach to building a house. Our library schools turn out craftspersons and contractors but not architects."⁽¹⁾

The U.K. Library and Information Research Group⁽²⁾ has discussed the gap between researchers and practitioners. Closer to my own home, the Library Board of Western Australia⁽³⁾ declared :

"In higher degree programmes ... there are two objectives that are often in conflict. The extension of professional expertise and the development of an ability to pursue independent reflective inquiry are not the same thing. The structure of a Master's programme will depend upon which of these objectives is given the greater weighting. In many areas of librarianship the extension of professional competence can be attained, by the interested librarian, from work experience. The methodology of research and problem solving is, however, far less likely to be assimilated on the job."

The relationship between Theory and Practice might be represented diagrammatically as in Figure 1. Practice is a source of primary data for hypothesis-formulation, and is also the potential field of application. Theory is thus the basis for decision-making, and it is the questions raised by decision-making which can enrich hypothesis formulation. There can also be two intermediaries between them : Education (where principles are instilled into potential practitioners) and in-house library research which should be the mechanism by which new theoretical developments can enter practice, and by which questions beyond the scope of one system can be referred to the theoretician. If this type of research is an interface between Theory and Practice, it becomes an important influence. Theoretical issues are transferred to the practical environment through the literature, which contains both theoretical dispute and reports of findings. User Studies illustrates the communication problems caused for in-house library research by the conflicting objectives of reports in the literature.

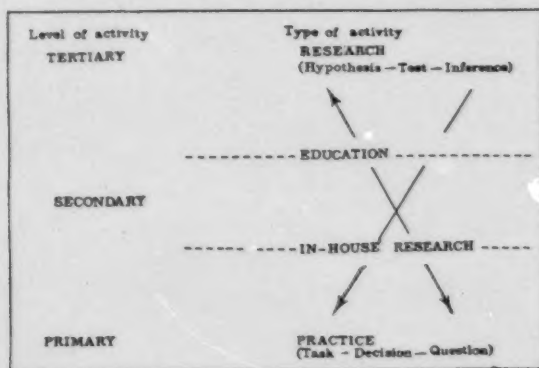


Figure 1: Relationships between Theory and Practice

The Literature and Methods of User Studies

In 1978 Vickery⁽⁴⁾ proposed a detailed model of the scope of Information Science (see Figure 2). He describes it as a "social process ... [of] people transferring information via documentary systems" This emphasis upon information as a social phenomenon, which logically demands research methods from the Behavioural Sciences. In the literature of User Studies between 1975 and 1980 there developed a marked shift, in the preference of researchers, from quantitative to qualitative methods.

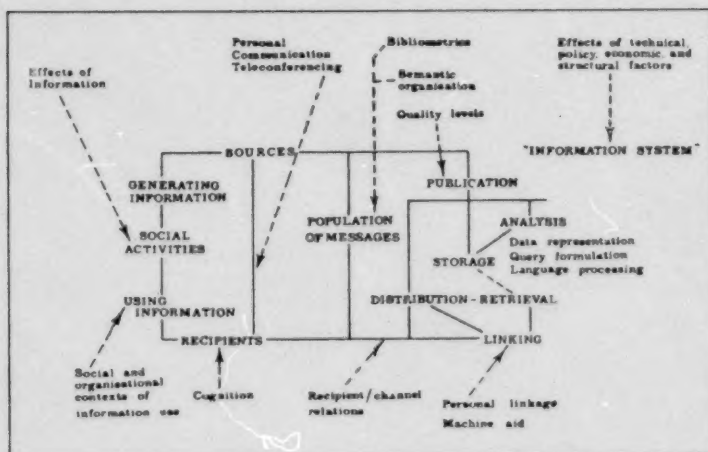


Figure 2: Overview of the domain of information research and its central problems

User Studies have been dominated by the use of social survey methods, although in 1975 Brittain⁽⁵⁾ called for methodological compatibility so that results could be cumulated as the basis for hypothesis-formulation;

in the same year Fabisoff and Ely⁽⁶⁾ attempted to synthesise the findings of studies of information needs into guidelines for system design.

This apparent stability was reinforced when in 1975 the Centre for Research on User Studies was established "to improve methodology and value of results⁽⁷⁾" by providing a national focus "for research, education, advice and reference information". All seemed set for a period of consolidation. But by 1979, Brittain⁽⁸⁾ was to emphasise the non-cumulative nature of Behavioural Science Research (the source of the methods of User Studies), Allred⁽⁹⁾ was to voice his suspicion of the social and ethical implications of Operations Research-based methods of service measurement, while a BLR&DD-funded symposium examined the "new wave" of qualitative methods being used by information researchers. This shift may be more specifically traced over the IRFIS papers, while a balanced account of the current position was recently outlined by Wilson⁽¹⁰⁾.

The cause of this shift is to some extent disillusion with the predictive value of probabilistic statements about human behaviour which is too fluid and intangible to be described so that observations can be cumulated in the same way as in the Natural Sciences. Although despite this adjustment in the theoretical literature, the bulk of user studies

conducted over this period continued to use empirical social survey methods⁽¹¹⁾.

This trend probably reflects the popular misunderstanding of User Studies - that we are simply inquiring what users think or feel about information systems. In fact, as we saw in Figure 1, we are examining a complex communication system, no matter how microscopically we focus onto it, and while the methods of the Behavioural Sciences are imperative, the classic social survey may be too blunt an instrument. In the literature of this period we might compare Lancaster's⁽¹²⁾ review of the methods of measurement and evaluation in information systems with Chen's⁽¹³⁾ unresolved debate over quantitative methods. There she contrasts the different needs of the Library Administrator, prevented by system parameters from considering much beyond internal efficiency, with those of the independent researchers, concerned with more theoretical problems. In User Studies the fundamental problem is what happens when book (information package) and user meet. The interaction is complex, and there have been various attempts to define the basic concepts⁽¹⁴⁻¹⁶⁾. Clearly these concepts are of central concern to other disciplines, and while some theoretical contributions to the User Studies literature are related to theoretical arguments in other disciplines, their value for us must lie in the tools they provide for observing and analysing the behaviour of users of information in order to improve this process of communication.

Conflicting objectives of user studies

Because of the healthy debate about the propriety of alternative methods, the objectives of research reported in literature can differ widely. We can see four main dichotomies in research objective.

1. Pure versus Applied research.
2. Theoretical versus Empirical research
3. Quantitative versus Qualitative research
4. Research into use versus usage

These dichotomies will now be discussed and illustrated as an aid to the categorisation of user studies.

The distinction between "use" and "usage" is fundamental. Use occurs inside people's heads, and cannot be directly observed, even by the attachment of instruments to the human brain (annotation 1). The dialogue between Belkin and Pratt⁽¹⁷⁻²⁰⁾ shows how alternative definitions of a concept such as "information" can yield quite different experimental models. But the in-house library researcher has more immediate access to usage, indirectly as a consequence of use. Usage data might be in the form of book disturbance⁽²¹⁾, circulation records⁽²²⁾ or citations⁽²³⁾, from which he could infer use. But as Borkowski⁽²⁴⁾ shows, circulation data is an objective and indirect

correlative which cannot describe all types of use. So the object of research which is most amenable to in-house library research is in fact one which is not the most useful. This is a cause of much dissatisfaction with individual user studies.

As we have just seen, there is a number of theoretical problems in User Studies, and much research is undertaken to try to solve them. But theoretical studies are felt to be of no practical help by the Library manager, while the theoretician bemoans the methodological impurity of pragmatic studies. Again, the conflicting objectives of the theoretician and the practitioner lead to dissatisfaction with the report. Another dichotomy exists between pure and applied research. For example, many User Studies are undertaken through the manager's needs for system-performance data. Such studies will be empirical. Nevertheless, empirical studies need to depend upon a body of theory, which Scientific Method tells us should grow by promotion of hypothesis to theorem through empirical validation. Many user studies are undertaken ostensibly as part of this theoretical process. So Line⁽²⁵⁾ might be classed as a theoretical contribution, since it sought to test certain hypotheses about service demand through experiment, rather than as an empirical exercise; that particular information service would not have existed other than to generate data for hypothesis testing. The distinction is slightly fuzzy, however, because the nature of support for information research makes it difficult to declare such a clear-cut distinction between theory and practice. Research in User Studies is not generally

undertaken without reference to the practical utility of its results, but the underlying orientation of much work may be seen as towards one of these two poles.

Finally, there is the dichotomy between Quantitative and Qualitative methods referred to earlier. The problem here is that what methodologists refer to as a qualitative method may be confused with what in the practitioner's world may be called "judgement". In other words, qualitative methods promise the academic researcher the possibility of gaining new insights into phenomena not previously examinable in their entirety while they may promise the practitioner freedom from the (now) unnecessary rigours of quantitative methods.

In Figure 3 we can see how these dichotomous objectives are reflected in two typical but hypothetical user studies. In the diagram, the four dichotomies overlap, to show how they are present in all studies. Hypothetical Study I is an attempt to test the hypothesis that students in different academic disciplines are to different extents dependent upon the Library. This may be performed empirically by statistical analysis of circulation records. In the diagram we can see which polarities are reflected in this study. On the other hand, in Hypothetical Study II, the hypothesis that the differing needs of students in academic departments for the same material requires variable loan periods, might best be tested by qualitative research into people, collecting data which

is compared with theories derived from research in other disciplines, and which may not be intended to form the basis for immediate action. Again, the shaded area indicates the polarity of this study.

It is, of course, easy to see how such studies could be written up in ways which would conceal their true objective, and thus make it unclear for which audience they were intended. Such a model explains much of the apparent difficulty in forming general conclusions from a diverse corpus of literature with few goals in common and even less in compatible or comparable results. It also enables us to sort out many misconceptions about the expectations of user studies: in order to provide usable answers, we may have to pose our questions less simplistically, and look at issues (such as inter-library lending, circulation figures) rather than user groups (historians, the General Public, Industry).

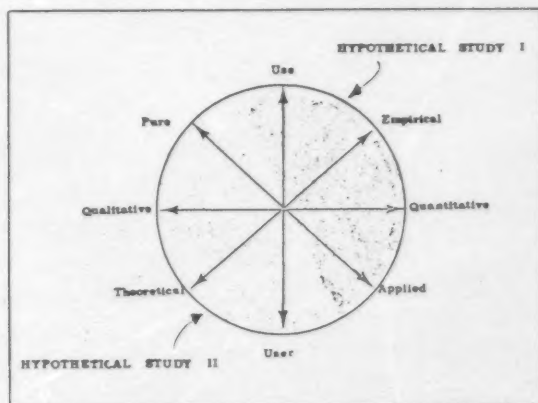


Figure 3 Dichotomies between Objects and
Methods of User Study

Debate about the relative propriety of different methods, can lead to the false belief that sociological investigation follows inherently different logical processes from those used in the Natural Sciences. To assert that different approaches yield data requiring different forms of analysis is not to justify abandoning the rules of logic. While the method of achieving academic consensus may differ, the verification of statements, and thus the manner of justifying them, should not, since the rules of inference transcend individual disciplines, even if the bases do not.

Thus we accept that many sociological hypotheses can only be tested initially by qualitative methods, and it may well be that we are going through an extended period of concept formulation requiring such approaches; but at some point all hypotheses, if they are not to remain unsubstantiated assertion, derive utility from empirical verification, where confidence relies upon success in prediction. There is good evidence to believe that the nature of behavioural data requires reformulation of much probabilistic argument, but not for their complete abandonment. Qualitative methods are appropriate to an early stage in the research process; they are not an independent method of enquiry.

Traditional social survey methods are creaking under the weight of cumulated, but non-conclusive, data. Skelton⁽²⁶⁾ showed empirically

that survey data was noncumulative, and Brittain⁽⁸⁾ has recently argued eloquently that the disorderly nature of data prevents the Behavioural Scientist from proceeding in the same orderly manner as the Natural Scientist. He shows how the sum of observed parts does not equal the living whole. Reluctance to use quantitative methods can also be due to a natural squeamishness at the use of mathematical descriptions of human beings. On the one hand, perhaps, there is a feeling that this is an affront to the human spirit, and on the other, suspicion of the motives of those who seek to predict human behaviour. It is naive to pretend that such issues are not at large even in an apparently innocuous area as User Studies.

The search for a homeopathic rather than aleopathic approach has led to increasing interest in the so-called qualitative techniques. Harris⁽²⁷⁾ has explained how "illuminative evaluation" might be used in library research, while Stenhouse⁽²⁸⁾ put his own long-gestated ideas into practice. The claim is that the epistemological problems of observing human behaviour can be overcome by the adoption of impressionistic (in extreme forms, "novelistic") techniques. The consequence is not only that data is often complex and difficult to analyse or replicate, and takes a long time to collect, but that the samples are so small that there is no statistical basis for generalisation. It is tempting to see, then, two schools: a Quantitative School, committed to a sort of "scientism" based on inhuman statistical techniques, opposed to a Qualitative School, defending the uniqueness of

human beings by intuitive techniques. This is, however, oversimplistic. In reality, those who use quantitative techniques would insist that the hypotheses these techniques support have been developed perhaps intuitively on the basis of in-depth study of small samples, while those who use qualitative techniques clearly have a belief in certainty, and thus in probability. We may therefore be grateful to the emphasis recently upon this earlier stage in the research process for its potential contribution to the necessary re-consideration of our existing hypotheses. We should not forget, however, that both types of method have to be used rigorously, and this implies labour, and thus cost. If we argue for qualitative research as a previously attenuated stage in research, we may be lengthening the process beyond the resources of in-house library researchers. This group may still be forced to choose, between affordable methods, and any suggestion that qualitative methods are inconclusive will reinforce their perhaps natural predilection for quantitative methods. We shall now, therefore, consider the relative costs of these methods before considering their relative appropriateness to in-house library research.

COSTS

The main reason for doing in-house library research is cost effectiveness: service-specific data can be collected as a by-product of normal operations by a researcher who is assumed to know and understand the system and its procedures better than an outsider. The in-house library researcher is supposed to know enough about methods to be able to

select appropriately, and design within pragmatic tolerances of precision. The costs of in-house library research becomes very important in deciding whether to mount a project, and in evaluating its findings. One of the advantages of quantitative methods is that research patrons are predisposed (once they have commissioned a project) to results based upon statistical inference: they appear objective, trustable and reliable. Much of the concern expressed amongst methodologists about quantitative methods concerns, in fact, their relative reliability compared with qualitative methods. Whilst this can be stimulating and confusing to the in-house library researcher, it may signal to the research patron reductions in cost.

When researchers seek funds, they have to quantify the resources they need in monetary terms. Usually these are dominated by staff costs, and the Australian Research Grants Scheme, for example, is typical in trying to confine these to "expert" or "professional" staff. When a "monopolistic" funder such as the British Library Research and Development Department appears on the scene it is natural to expect an accumulation of experience generating standards for the allocation of funds. In practice, however, it is difficult to define unit costs for research, in the way that it is possible for library operations, and thus to compare the cost-effectiveness of different methods of conducting research. One reason why this assumes major importance is that researchers used to selecting methods pragmatically, rather than ideologically, might expect to use cost-effectiveness as a yardstick. In practice, it is difficult to apply such a standard. So in the debate

about the propriety of qualitative or quantitative methods, we have so far seen little attempt to relate the unit costs of using the two types of method to the utility of the information gained, and no assessment of the propriety of using cost-effectiveness as a criterion for method-selection. In other words, in-house library researchers have little evidence that either of these two methodological "schools" is significantly cheaper than the other, or that the results are inherently more useful. Such information is critical to researchers working in an environment not dominated by the British Library Research and Development Department. These groups desperately need criteria by which to select one method against another. In an attempt to evaluate the relative cost-effectiveness of alternative methods, let us consider the unit-costs (annotation 3), and utility of results, of two "classic" studies conducted at WAIT during the last two years.

It is difficult to find two studies which are exactly comparable. Ideally they should have been directed to the same type of question (or the methods would not be "interchangeable"), conducted under comparable conditions, and have accessible cost data. Presumably a large funder such as BLR&DD might have projects satisfying all three criteria. A programme of in-house library research would find it difficult to satisfy all three. However, two studies, which come fairly close from the W.A.I.T. Library are a study of reciprocal use amongst academic libraries in Perth and a study of W.A.I.T. Faculty attitudes towards the Library.

They are both "user studies", each use interviews and both had ascertainable cost data. However, the former project was unashamedly quantitative in approach, while the latter project was an attempt to use the techniques of Illuminative Evaluation described by Stenhouse⁽³⁰⁾.

CASE-STUDY A

Perth (WA) is a relatively small, isolated community. The (combined) stock of monographs and serial titles of its tertiary academic libraries is contained in four library systems. A Reciprocal Borrowers' Scheme⁽²⁹⁾ and Network Committee (annotation 4) represent attempts to optimise access to this stock. In 1981 the Commonwealth Tertiary Education Commission⁽³⁰⁾ recommended more cooperation between tertiary academic libraries, and in an attempt to quantify the extent of existing sharing Project SOUL⁽³¹⁾ was conceived. This study attempted to quantify the extent to which members of each academic institution used the libraries of the others. The method selected was that of the classic quantitative social survey. A sample of observation times was drawn for each institution. A short interview was conducted across a sample of respondents, and their responses subjected to statistical analysis. Ideologically it was required that assertions based upon interpretation of the researchers' observations should be "justified" by reference to the laws of probability. The reasons for this ideological position are necessarily obscure.

In attempting to calculate the unit-costs of this project, we have to consider some unusual features of the method, since they undoubtedly inflated the costs. Seventeen service points were nominated, and an attempt was made to conduct interviews at least once on each day of the week. These "days" were spread over three months (to allow for seasonal fluctuations in widely differing academic calendars); theoretically they might be telescoped to mirror "typical" weeks. Interviews were conducted over two hour periods. Two interviewers were assigned to each interview period. One interviewer contacted each person exiting from the library service point, interviewing only the foreigners. The other interviewer contacted a person exiting at predetermined intervals; this interviewer interviewed both foreigners and natives, and thus sampled the stream of people exiting. This interviewer's results were regarded as a benchmark against which the foreigners interviews could be measured.

Four thousand, seven hundred and twenty-eight interviews were completed in 299 hours of interviewing time. The interviewers were untrained staff, paid at casual rates. The total, real, estimated costs incurred by this project were A\$14530. These costs were inflated by the encoding. Moreover, it was decided for pragmatic reasons that it was more appropriate to "employ" in encoding the senior member of staff responsible for a project, rather than hire extra clerical staff. Equally pragmatically it was decided to employ the R & D Librarian for the computer analysis. Finally, WAIT is unusual in charging real money for its computing: the costs therefore represent a real cost to the

library, but may be compared with the Institute's perception that similar charges in the outside world would be ten times higher. It should be pointed out that all the calculations of costs in this paper, exclude indirect costs associated with the library buildings and accommodation.

In calculating the unit-costs, we have to ask what is the unit? The simplistic answer is that it is the interview. The unit cost of the interview is therefore A\$3.07. Against this view we have to place the importance attached to the individual question in the interview: the fact that 13 were selected suggests that they were regarded as separate and important pieces of information. When the question becomes the unit, the unit-cost falls dramatically to A\$0.24. This, however, quickly emerges as an equally simplistic strategy. Most of the answers yielded by the individual questions are meaningless without the kind of cross-tabulation in which SPSS excels, and whose significance non-parametric tests evaluate. The presence of 13 questions clearly presents a very large number indeed of potential permutations ($13!$), and offers the kind of fishing expedition against which the SPSS manual itself warns. But even assuming a modest number of potential cross-tabulations, clearly the interview question is not a simple unit. If we used the number of potentially meaningful cross tabulations, we can further reduce the unit costs even more dramatically.

The meaningfulness of cross-tabulations is a product of the following factors :

- a) The statistical significance of association.
- b) Intuitive judgement of the utility of statistical significance.
- c) The researcher's imagination.
- d) The simplicity of the data cross-tabulation.

It has also to be said at this point that the utility of the data is also hard to determine or quantify. The study was undoubtedly intended as a general, possibly political, contribution to national and local debates concerning the sharing of library resources.

Attempts to trace the diffusion of innovation show how difficult it is to establish a general framework upon which might be based calculations of cost-effectiveness. It becomes, therefore, impossible to try to calculate the value of the utility of these results.

CASE STUDY B

The T.L. Robertson Library at W.A.I.T. is a centralised library service for the campus. It is experiencing most of the managerial problems

imposed by external socio-economic forces; the same problem is experienced by academic libraries in other OECD countries. As a contribution to decision-making in this environment, it was decided to investigate Faculty and student attitudes to the library.

Considerable debate occurred amongst library management as to the best method of conducting such a study. The issues in this debate were principally objectivity (and thus credibility) and scale (cost). The consequence was that a senior member of staff in the Institute's library school was commissioned to do the study, and resources were supplied as requested by the library (staff and consumables). Objectivity was preserved by giving the research director complete freedom to operate as he chose, and by giving the research assistant his own office in the library school rather than the library.

The seriousness with which the researchers took their academic freedom is exemplified by their swift replacement of the classic social survey recommended in the research patron's brief by the methods of illuminative evaluation as described by Stenhouse⁽³²⁾. The reason for their choice was quite plain. A study of such complexity clearly required an initial period of in-depth examination of the problem as the basis for subsequent hypothesis-formulation and probabilistic testing. Thus the researchers readily agreed to develop from their findings a questionnaire which could be administered to a sample of the population. This methodological

courage incidentally, might be compared with the attitudes which clients of a monopolistic funder might bring to such methodological crises.

The researchers decided to use Themes Analysis⁽³³⁾, a technique (derived from Market Research) whereby panel discussions identified themes for subsequent investigation, as a preparation for the type of interview the researchers believed Stenhouse had conceived. They then conducted 40 of these interviews, had them audio-typed, sanitised by the interviewees, re-typed, and archived. The principal researcher then indexed the interviews, in an attempt to identify perceptions and attitudes, which he recorded in a final report. All research materials were then accumulated in an archive which, with the report, was presented to the research patron.

Once again there is a number of uncertainties associated with the costs incurred by this stage of the project. For example, the level at which the research assistant was appointed reflected the marketability of the individual who became available, rather than a particular recruitment policy. The number of interviews conducted was equally a function of the time available (the duration of his contract of employment) than of any precise methodological calculation. Several cost elements (such as the audio-typing) are costed at the rate of the operative selected, rather than that of the optimal ideal source of such services. The total cost

was A\$17,920. The cost per interview is A\$448. This much higher cost is not surprising, since these interviews lasted on the average for 30 minutes, and were of necessity unstructured. An essential aspect of this method, however, as interpreted by these researchers, is the presentation alongside the research report of the research archive. In the case of this project the archive consisted principally of the sanitized transcripts of the interviews, indexed and cross-referenced to the report. It is arguable what now becomes the unit of costing. The products are of such obviously unequal size and importance that these are unsuitable. The number of access-points to themes was approximately 300, representing a much lower unit cost (\$59.73). This unit cost can be further reduced by considering the total number of occurrences of each access point, and even lower if we consider the theme-occurrences. A fundamentalist view, that the words in the sanitized texts of the transcripts of the interviews are the units of costing gained support from the theory underlying this method: the reporting researcher's interpretation of these words is the key element in this research method. A simple division of total cost by the number of word-tokens in the text, yields a meaninglessly small unit cost. However this is clearly too simplistic.

Determining the unit costs of in-depth interviews is not therefore easy. In summary the Qualitative study was 23 percent more expensive, overall, than the Quantitative study, and had a markedly higher unit cost (A\$448 against A\$3.07). It will be possible, methodologically, to argue for

different quantities for interview in both studies. But it is clear that such enhancement of reliability would be significantly more costly in the Qualitative study. The only reliable unit seems to be at the high level of the complete interview, although it seems to yield in these examples very uncompetitive results. In examining the utility of these results, we experience further problems. In the first place, although the results were the basis for the design of an instrument which was used, it is difficult to trace precisely the effects of the products of the first stage on the instruments of the second. Indeed, the unusual use of this method, as an elongated stage in the quantitative research process rather than a substitute for it, seems to yield little precise information about this usually unreported stage.

INFERENCES

The serious debate between the proponents of qualitative and quantitative methods is usually conducted on theoretical grounds. These grounds reflect perceived propriety of methods, which may be based on more general philosophical positions. It is arguable that this debate is of interest only to academic or centrally-funded researchers. In-house library researchers, it might be argued, are constrained by the sample sizes and the resources available to them, and the methodological prejudices of the audience for whom they are reporting their findings. If such a group has a choice between methods, then one might expect their choice to be governed by practical considerations such as cost. The

purpose of this paper is to emphasize the difficulties in establishing the relative unit-costs of such research. However, it is also true that the bulk of the costs are attributable to the salaries of the staff assigned rather than hired.

The cross which in-house and government-funded researchers have to bear is justification of their research. Incalculable quantities of in-house, mission-oriented investigations (which presumably result in management decisions), are not reported on the basis that they are not "research". Formal research is, therefore, presented in the distortion of a chiaroscuro born of the suppression of the former category: having limited practical utility its failure to be fully utilised becomes its condemnation. When publicly-accountable institutions put up the money, it becomes harder to justify doing this kind of research. Without mechanisms for tracing the use of research (let alone the cost-effectiveness of such a use), it becomes impossible to see any pragmatic basis for the selection of methods.

Until such criteria become available, the basis for selection might be any one, or combination, of the following :

- a) Philosophical predilection (if the researcher is equipped to rationalise them)

- b) The methods one has used in the past (if the researcher has previous research experience)
- c) The methods used by previous research into the problem (if it has been researched, reports made public, a methodological consensus established, and the researcher has read the literature).

It seems then that theoretical researchers in User Studies are concerned that qualitative methods may yield more insights than quantitative methods while the in-house library researcher may lean more towards quantitative methods for pragmatic reasons. Therefore there is a methodological divide between theoretical and in-house library research which threatens to inhibit communication: theoreticians interested in qualitative methods are liable to be starved of data, at least in the short term!

This author concludes sadly that there is an inevitable and irreconcilable gap between academic and in-house library research, caused by the restrictions imposed by resources upon what the in-house library researcher can try to observe. In the area of User Studies there are strict limits on how close the in-house library researcher can come to observing use.

In Figure 4, we see use ranging from use inside the library to use outside it: intuitively we can accept that these are not necessarily dichotomous activities. In this diagram, we see calibrated in the continuum in terms of information usage. Further, we can see five aspects of this usage:

1. The extent to which use is made of documentary material.
2. The relative proportions of library-owned and personally-owned material.
3. The degree of formality in the use.
4. The extent to which the literature tells us that it is a preferred type of use.
5. The extent to which it is an observable form of use.

This somewhat condensed and personal account is intended to show in a general way that the in-house library researcher is limited in the extent to which he can observe the phenomenon of use. Thus the line CD is the general cut-off point beyond which the phenomena are totally unobservable. Perhaps, Wilson's study of information transfer events amongst local government officials⁽³⁴⁾ represents the furthest the in-house library researcher can realistically go in observing non-documentary information transfer. But, the in-house library researcher cannot attempt to get this close to personal communication, and so the line AB represents in a general way the limits to his horizons.

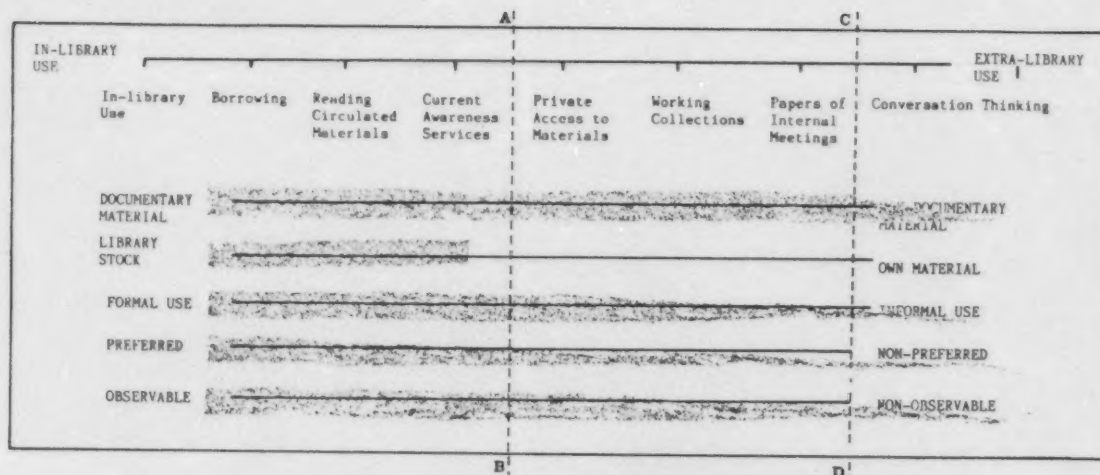


Figure 4: Observability of different types of information use

Key: Extent of applicability to site of information use.

Notes:

Information use can be characterised by its dependence upon the Library; this dependence is represented by a continuum of use occurring to varying extents inside or outside the Library.

Different inferences may be drawn about:-

1. The extent to which the information use is dependent upon documents.
2. Ownership of the material.
3. Degree of formality of the communication process.
4. Degree of user preference for material.
5. Degree of observability of use.

The dotted line AB indicates the maximum likely penetration of library-based research into the organizational communication process, while the line CD indicates the maximum possible penetration.

The inference which this researcher draws is that the restricted landscape permitted to the in-house researcher is more amenable to quantitative than qualitative methods. He is cut off from personal contact with the user, particularly in the use-behaviours and contexts most amenable to qualitative research. The data most readily available is more reliably treated quantitatively. His reporting structures are predisposed to receiving quantitative data.

If such inferences are true, then in-house library research cannot readily transfer qualitative methods from academic research, even though the rigors of quantitative research may have been compromised by their use in small-scale research*. Consequently, academic researchers predisposed to qualitative methods may be deprived of primary data, as well as access to field-tests of their methods. It remains to be seen if this separation leads to non-implementation of the useful products of qualitative methods.

This paper has attempted to illustrate the general problem of the relationship between academic and in-house library research by looking at a particular methodological problem in User Studies, namely the relative cost of different methods. This analysis of costs suggests that while the total costs seem to be very similar, their relative credibility may be markedly unequal because of their different rates of diffusion amongst the profession. Despite the appearance of articles on quantitative

methods in the professional literature^(27,35), there is still something of a credibility gap amongst libraries educated in quantitative methods. Recently, Harris⁽³⁰⁾ urged the need to make librarians "so research-conscious that when they are faced with a problem they think automatically of undertaking research". A step in this direction is the increasing number of library posts dedicated to research in USA, UK and Australia. A next logical step would be to link such posts more directly with centres of theoretical expertise, in library schools and/or research units, thus completing the chain of diffusion proposed in Figure 1.

ANNOTATIONS

1. The American poet John Forbes recently observed "you can't be ironic about things you see on television because what you are seeing are just patterns of electrons, only real inside your head where they are given meaning". [Australian] National Times No. 647 (24-30 June 1983) p.32.
2. Costs are expressed in Australian dollars calculated in 1982 Australian prices. Readers more familiar with other currencies should initially convert them at the average exchange rate prevalent in that year, but should also be warned that prices may not be directly convertible.
3. The Western Australian Network Committee was formed in 1980 to plan a computer-based library network for the State.
4. For example, Barnes (a delegate to IRFIS-4) sparked off a debate about the reliability of the chi-square test when applied to the small quantities of observations collected during her own exploratory research. This debate occurred in Library and Information Research News.

REFERENCES

1. WHITE, H.S. Research: the lumber yard approach. *Library Journal* (15.12.1981) 2347.
2. Are Libraries getting the research they deserve? *Library and Information Research News* 3,11 (Sept. 1980) 3-.
3. LIBRARY BOARD OF WESTERN AUSTRALIA Statement on education for research in Librarianship. IN:- WAIT Submission for accreditation of Graduate Diploma in Advanced Library and Information Studies Part II. WAIT, 1982.
4. VICKERY, B.C. and BAXTER, P. Workshop on basic information research BLR&D Report no 5500.
5. BRITTAI, J.M. Information needs and application of the results of user study IN:- DEBONS, A. and CAMERON, W.J. Perspectives in information science. Leyden, Noordhoff, 1975.
6. FABISOFF, S.G. and ELY, D.P. Information and information needs. *Information Reports and Bibliographies* 5,5 (1976)2-16.
7. CRUS. Background Note [1975].
8. BRITTAI, J.M. Information services and the structure of knowledge in the social sciences. *International Social Science Journal* 31,4 (1979)711-728.
9. ALLRED, J. The measurement of Library services; an appraisal of current problems and possibilities. Bradford (W. Yorks), MLCB Publications, 1980.
10. WILSON, T.D. On Information Science and the Social Sciences. *Social Science Information Studies*. 1,1, (1980)5-13.
11. FORD, G. User studies: an introductory guide and select bibliography. BLR&D Report No. 5375.
12. LANCASTER, F.W. The measurement and evaluation of Library services. Wash. (DC) Information Resources Press, 1977).
13. CHEN, C.C. Quantitative measurement and dynamic Library service. Phoenix (Az.), Oryx Press, 1978.
14. LINE, M.B. Draft definitions: information and library needs, wants, demands and uses. *Aslib Proceedings* 26,2(1974)87.
15. ROBERTS, N. Draft definitions: information and library needs, wants, demands and uses. *Aslib Proceeding* 27,7(1975)308-313.
16. WILSON, T.D. On user studies and information needs. *Journal of Documentation* 1,2(1981)3-15.

17. PRATT, A.D. The information and the image. *Libri* 27,3 (1977) 204-220.
18. BELKIN, N.J. Information concepts for information science. *Journal of Documentation* 34,1(1978)55-85.
19. PRATT, A.D. Information concepts. *Journal of Documentation* 34,3(1978)242-243.
20. BELKIN, N.J. Reply *Journal of Documentation* 34,3(1978)245-246.
21. Project for evaluating the benefits from university libraries: final report OSTI 5056.
22. KENT, A. et al A cost-benefit model of some critical Library operations in terms of use of materials: final report. New York, Marcel Dekker, 1979.
23. GARFIELD, E. Citation indexing - its theory and application in Science, Technology, and Humanities. New York, Wiley, 1979.
24. BORKOWSKI, C and MACLEOD M.J. The implications of some recent studies of Library use. *Scholarly Publishing* 11,1(1979)3-24.
25. LINE, M.B. et al Experimental information service in the Social Sciences BLR&D Report No. 5085.
26. SKELTON, B. Scientists and social scientists as information users: a comparison of results of science user studies with investigation into information requirements of the Social Sciences. *Journal of Librarianship* 5,2(April 1973)138-156.
27. HARRIS, C. Illumination evaluation of user Education programme. *Aslib Proceedings*. 10,29(1977)348-362.
28. STENHOUSE, L. The problems of standards in illuminative reserve. Paper presented at AGM Scottish Educational Research Association, Glasgow, Feb. 1979 [private communication].
29. HALE, B. Reciprocal use of academic Libraries. Paper presented to Library Association of Australia Biennial Conference, Adelaide [in publication].
30. [AUSTRALIA] COMMONWEALTH TERTIARY EDUCATION COMMISSION Report for the 1982-84 Triennium, Vol. 1, Part 1, Recommendation on guidelines. Canberra, AGPS, 1981, p155.
31. ALLEN, G.G. and EXON, F.C.A. Interim report on a study of the users of academic and research libraries in Western Australia Perth (W.A.), WAIT, Library, 1983.
32. LABBETT, B. The Library access and sixth form studies project. *CRUS NEWS* 11(1981)7-8.
33. SMITH, L. [private communication].

34. WILSON, T.D. et al Information needs and information services in local authority social services departments (Project INISS) BLR&D Report No. 5453.
35. HARRIS, K.G.E. Do Library schools really understand academic needs? Library Association Record 84,10(1982)351-352.
36. ODI, A. Creative research and theory building in Library and Information Sciences. College and Research Libraries 3,2(1982)312-319.

Environmental monitoring and information management
in County Planning Authorities

by

T.D. Wilson and I.M. Masser

University of Sheffield, Sheffield S10 2TN, UK

1. Introduction. This paper reports on the results of a census of County Planning Authorities (CPAs) in England and Wales using a mailed, self-completed questionnaire, intended to obtain information on the state of information management in these authorities and the relationship of information management to aspects of environmental monitoring.

By their very function CPAs are 'self-evaluating' organizations: the function of planning carries with it a necessity to discover something about the impact of plans when they are put into effect and to understand the nature of change in the environment for which plans are intended. In other words, a process of 'organizational learning' (Argyris & Schon 1978) must take place if the organization is to continue to fulfill the aims for which it is established. Studies such as those by Wildavsky (1972) point to the central position played by information management practices in organizational learning.

Despite the frequent re-iteration of the importance of information management, however, little is known about the relationship to organizational learning in the field of urban planning. Most models of the implementation process stress the interaction between policy and action and stress the evolutionary nature of the process (for example, Barratt & Fudge, 1981) and deal with information practices only in passing. Similarly, studies in the field of information management tend to be concerned either with narrow technical questions or with human factors aspects and

the literature on organizational impacts is much less extensive.

In the study reported here the focus of the investigation was 'monitoring' - that is, those activities of CPAs which relate to the impact of plans or the discovery of environmental information related to plan-making. In the planning literature considerable attention has been given to the monitoring function and models exist which relate organizational learning to data collection, information dissemination, and information handling (see, for example, Faludi's (1973) cybernetic model of the planning process).

The investigation was confined to CPAs for two main reasons:

- * because these authorities have taken a leading role in the implementation of monitoring systems because of their statutory responsibilities for the preparation of 'structure plans' (Cater, 1979; Van Meyel, 1979);
- * because of their emphasis on strategic monitoring rather than on a more narrowly conceived performance monitoring (Floyd, 1978);

2. The investigation. The investigation was carried out using a structured, self-completed questionnaire in which views were sought in terms of their polar extremes. Respondents were asked to select the extreme that corresponded most closely to their own situations.

There are clear disadvantages in such an approach: firstly, respondents have no opportunity to qualify their answers, other than in added comment or in a covering letter, and may find it impossible to categorise themselves according to the options presented in the questionnaire. Secondly, however well they are constructed, the polar extremes are unlikely to appear as mutually exclusive to all respondents because of differences in local practice. Thirdly, there is the danger of over-categorization, in that respondents may not be aware of the hypotheses

implicit in the categories. Despite these difficulties, however, it was felt that the approach was likely to be more satisfactory than a factual, data-collection exercise, particularly for a preliminary investigation.

3. Research hypotheses. A two-part questionnaire was sent to all County Planning Authorities in England and Wales with instructions that the section asking for contextual information on current planning activities should be filled in by the County Planning Officer or someone else with a broad overview of the activities of the Planning Authority as a whole, while the section on attitudes to information management should be answered by the officer most closely concerned with this task. In addition to the questions related to the hypotheses described below respondents were asked for supplementary information on their information activities and for their views on the likely impact of current developments in information technology.

3.* Planning hypotheses: these hypotheses and the associated results have been presented in another paper by the same authors (Masser & Wilson, 1983) and will not be reviewed in detail here.

The planning hypotheses related to planning priorities, the status of monitoring within the Authorities, and various aspects of their approach to monitoring. They can be stated briefly, in null form, as follows:

PH1: there will be no differences in planning priorities relative to the date of Structure Plan approval.

PH2: there will be no differences in the extent to which monitoring activities are integrated with other departmental functions relative to differences in planning priorities.

PH3: there will be no differences in the main purpose of monitoring (as either reviewing progress or identifying new choices and problems) relative to differences in planning priorities.

PH4: there will be no differences in the extent to which reports on monitoring are directed towards internal or external consumers relative to differences in planning priorities.

3.2 Information management hypotheses: It was anticipated that information management practices would vary among the respondents in ways related to their overall planning priorities and to the status given to monitoring. Consequently, the following hypotheses were tested (again given in null form):

IH1: there will be no differences in the scope of information collection relative to differences in planning priorities.

IH2: there will be no differences in the extent of use of computerised versus manual information systems relative to differences in planning priorities.

IH3: there will be no differences in the extent of use of computerised versus manual information systems relative to the variation in state of development of monitoring activities.

IH4: there will be no differences in the extent to which special information systems have been developed for monitoring relative to differences in planning priorities.

IH5: there will be no differences in the extent to which special information systems have been developed for monitoring relative to differences in the state of development of monitoring activities.

IH6: there will be no differences in the extent to which needs are perceived for systems which deal with 'hard', quantitative information rather than 'soft', qualitative information relative to differences in planning priorities.

IH7: there will be no differences in the extent to which needs are

perceived for systems which deal with 'hard', quantitative information rather than 'soft', qualitative information relative to differences in the state of development of monitoring activities.

IH8: there will be no differences in reporting procedures in monitoring relative to differences in planning priorities.

IH9: there will be no differences in reporting procedures relative to differences in the state of development of monitoring activities.

It was also felt that there would be relationships among the information management variables resulting from variations in practice and, therefore, the following hypotheses were also tested:

IH10: there will be no differences in perceived needs for systems which manage 'hard' information relative to differences in the scope of information collection.

IH11: there will be no differences in reporting procedures in monitoring relative to differences in perceived needs for systems which manage 'hard' information.

Respondents were also asked to rank the perceived importance of three operational problems in information management: the lack of relevant information; problems in the retrieval, processing and evaluation of information; and lack of effective channels of communication; and two further hypotheses were tested in relation to these problems:

IH12: there will be no differences in the perceived importance of operational problems relative to differences in the scope of information collection.

IH13: there will be no differences in the perceived importance of operational problems relative to differences in perceived needs for systems to manage 'hard' information.

4. Results. Introduction. Questionnaires were sent to the 54 County Planning Authorities in England and Wales in mid-November 1982; a follow-up letter, with an additional copy of the questionnaire, was sent to those who had not responded to the initial request at the end of January 1983. The response was excellent with replies being received from 51 authorities, or 94.4%, of whom only three declined to answer the questionnaire. The remaining 48 responses (88.9%) contained sufficient information for analysis and, in many cases, gave added comments and additional material designed to assist in the evaluation of their responses. On the basis of response rate alone, therefore, it can be claimed that the questionnaire touched upon a theme perceived to be of importance to the Planning Authorities and this was borne out by statements in covering letters to the effect that respondents were eager to see the results.

As expected, a number of respondents had problems in defining their activities in terms of the categories included in the questionnaire. In a number of cases tied ranks were given and respondents devised their own categories. These problems were most serious in the question on planning priorities where 7 of the 48 respondents were unable to distinguish among priorities and several more ranked two or more priorities equally.

With these exceptions, however, the vast majority of respondents appear to have had little difficulty in completing the questionnaire. Even so, the findings of the survey must be treated with some caution as it is difficult to assess the extent to which arbitrary choices may have been made. The questionnaire, as noted earlier, was explicitly directed towards different parts of the organization and there is evidence to suggest (particularly among those who responded after follow-up) inconsistent and apparently contradictory views between these parts of the

organization.

The interpretation of the responses with respect to the hypotheses listed above is fraught with difficulties: some of the problems have been noted above and include tied ranking of some issues by some respondents, a degree of non-response, and the difficulties experienced by some respondents in answering some of the questions. In most cases the Chi-squared results reported for the hypotheses are lower than the level usually required for acceptance: however, given that a census was conducted, rather than a survey, we have chosen to take a level of $p=0.10$ as indicating 'some support', $p=0.05$ as 'support' and $p<0.05$ as 'strong support' for rejecting the null hypothesis.

In the light of the difficulties and the exploratory nature of the work, the following comments on the hypotheses should be regarded as tentative and open to clarification on the basis of further work.

4. The planning hypotheses. Of the four hypotheses under this heading two found some support in the results: PH1 and PH3. It is clear from Tables 1 that the date of Structure Plan approval has some relationship to the main perceived priority of the CPAs. For the purpose of analysis authorities were divided into two categories: those that gave top priority either to the control of new development or to the maintenance of the existing landscape, and those that regarded the promotion of economic growth or 'securing adequate provision for less-privileged groups' as being either their top or equal first priority. More than three quarters of the post-1980 Structure Plan approvals are in respect of authorities which fall into the second category.

Table 2 shows, also, that authorities that fall into the second group are more likely to see the function of monitoring as aiding the identification of new problems and choices, rather than as helping the review of

progress towards policy implementation. One would expect this latter finding to be of relevance to information management practices. The value for chi-squared is not significant even within the looser constraints adopted here but the figure of 14 in cell D of the table is, nonetheless, very suggestive of a relationship.

Main planning priority	Structure plan approval date	
	Before 1980	After 1980
New development/Existing environment	9	5
Economic growth/Less privileged	9	18

Table 1 (Chi-squared = 3.59 with 2df p.<0.10)

Main function of monitoring	Main priority	
	Development	Economic growth
To review progress	8	9
To identify problems	6	14

Table 2 (Chi-squared = 1.14 with 1df p.>0.20 but <.30)

In relation to PH2 and PH4 the results suggested that authorities generally regarded the monitoring function as being integrated with other functions of the Department and as being directed towards the provision of timely information rather than as increasing general awareness. This latter result is a little surprising as it was anticipated that the 'post-1980 approval' authorities, oriented towards economic growth and provision for less privileged groups, would be more inclined to look for monitoring systems which aided general awareness.

4.2 The information management hypotheses. Of the 13 information management hypotheses 7 found some support in the results: IM 2, 7, 9, 10, 11, 12, and 13.

Three of the hypotheses relating information management issues to

monitoring are supported by the data (if only weakly in some cases).

First, IH2 linking type of information system to planning priorities: the relationship is rather weak but there is a strong probability that authorities with priorities for economic growth and support for less-privileged groups (which also tend to be those with post-1980 Structure Plan approval) tend to have adopted computerized systems more readily.

Planning priorities	Type of information systems	
	Manual	Computer
'Traditional'	4	10
'Other'	2	23

Table 3 (Chi-squared = 2.92 with 1df p.<0.10)

Secondly, IH7 links perceived system needs with the state of development of monitoring: here there was just a suggestion here that authorities with 'evolving' practices were more likely to perceive their needs to be for systems to handle 'soft' information. Authorities with established practices were more likely to perceive a need for 'hard' data.

Information system needs	State of development of monitoring	
	Established	Evolving
'Hard data' systems	11	11
'Soft information' systems	5	14

Table 4 (Chi-squared = 2.40 with 1df p.>0.10 but <0.20)

Thirdly, the state of development of monitoring is associated in IH9 with reporting procedures. Authorities with established procedures are slightly more likely to produce regular reports for committees and officers, whereas those whose procedures are still evolving are much more likely (12 authorities to 3) to report on an *ad hoc* basis. Table 5 shows the strongest basis for the rejection of any of the null hypotheses in this group and it may be that the 'communication factor' inherent in

this relationship has strong associations with aspects of organizational structure and functioning generally.

Reporting procedures	State of development of monitoring	
	Established	Evolving
Regular reports	17	14
Ad hoc response	3	12

Table 5 (Chi-squared = 4.99 with 1df p.<0.05)

The remaining four hypotheses with some degree of support all link one aspect of information management with another. The distinction between systems for 'hard data' and those for 'soft information' figure in three of the four, and information acquisition practices (defined as either 'comprehensive' in scope or 'restricted') also figure in two of the four.

On the basis of Tables 6 to 9 one can summarize briefly that authorities which restrict their information acquisition activities to material relevant to major, immediate planning issues are much more likely to feel a need for systems to handle 'hard' data. They are also more likely to perceive a lack of relevant information as their principal operational problem in information management.

Information system needs	Approach to data collection	
	Restricted	Comprehensive
'Hard data' systems	9	3
'Soft information' systems	7	12

Table 6 (Chi-squared = 10.78 with 1df p.<0.01)

Authorities with 'comprehensive' acquisition policies are more likely to feel a need for systems to handle 'soft' information and are more evenly divided over the operational problems with one fewer choosing 'lack

of relevant information' than the other two categories. This pattern is repeated in Table 9 revealing the link between 'hard' data and lack of relevant information.

Reporting procedures	Information system needs	
	'Hard data'	'Soft information'
Regular reports	17	10
Ad hoc response	6	9

Table 7 (Chi-squared = 2.05 with 1df p.>0.10 but <0.20)

Operational problems	Approach to data collection	
	Restricted	Comprehensive
Lack of relevant information	15	5
Retrieval, processing, evaln.	8	6
Channels of communication	3	6

Table 8 (Chi-squared = 4.60 with 2df p.0.10)

Operational problems	Information system needs	
	'Hard data'	'Soft information'
Lack of relevant information	14	4
Retrieval, processing, evaln.	5	7
Channels of communication	2	7

Table 9 (Chi-squared = 7.79 with 2df p.<0.05)

5. Discussion. Two 'stereotypes' appear to emerge from this investigation. The basis for the stereotypes appears to be related to a division between those authorities that define information as 'hard data' and those that have a wider definition which includes 'qualitative' information. CPAs that adopt the former definition are likely to have a policy of restricted information acquisition and regular reporting to committees and others. They are also more likely to perceive there to be a lack of

relevant information. In relation to the planning characteristics these authorities tend to give highest priority to traditional land use matters.

CPAs that adopt the more 'general' definition of information are more likely to have a policy of comprehensive information acquisition and to make use of ad hoc reporting practices. Authorities falling into this stereotype are more likely to regard information retrieval, processing and evaluation or lack of effective channels of communication as the most serious information management problem. There is some connection between these characteristics and the more general planning characteristic of priority for economic and social issues.

We would not wish to over-emphasise the existence of these two stereotypes because the relationships among the component elements of the two sets are not always strongly marked. This is particularly evident in the case of the relationship between the responses to the question in the planning section on the function of monitoring and that in the information management section on the practices of information acquisition. On a priori grounds it might have been expected that there would be a reasonably close correspondence between authorities that regard the main function of monitoring as being to review progress towards plan-implementation and those that have adopted a policy of restricted information acquisition. One might also expect that those viewing the main function of monitoring as being to identify new problems and choices would also be those adopting a policy of comprehensive information acquisition. The findings, however, show no clear pattern at all in this respect.

The rather weak relationship between information management variables and planning variables leads to the formulation of a further hypothesis - that because information management is seen as a technical task undertaken by specialists of one kind or another (and increasingly by specialists in

computing), it is to some extent divorced from organizational characteristics. This is not to say that the subjects of the documents or data are unrelated to organizational functions but that the 'management' of the information-bearing media is sufficiently specialized and distinct not to be affected by organizational characteristics. This is clearly a possibility well worth further study.

6. Conclusions. The work reported here was of an exploratory nature and some of the findings are ambiguous in character. The limitations of a purely 'quantitative' approach to research in this area are revealed and we would argue that there is a need for further work to identify variables related both to planning and to information management which will show greater discriminatory power among CPAs. We would also argue that further work ought to take a 'case study' approach in which those authorities which show clearest evidence of following one 'stereotype' position or another are studied first so that the real nature of the distinguishing characteristics can be uncovered.

The possibility, noted above, that information management practices may not be closely related to other aspects of the organization is of some general significance and any further study should explore this issue in particular in greater depth.

References

- Argyris, C. & Schon, D. (1965) Organizational learning: a theory of action perspective. Reading, Mass.: Addison Wesley.
- Barratt, S. & Fudge, C. (1981) Policy and action: essays on the implementation of public policy. London: Methuen.
- Cater, E. (1979) Patterns of information use in planning: a study of Oxfordshire and Berkshire. Oxford: Oxford Polytechnic Department of Town Planning. (Working Paper 39)
- Faludi, A. (1973) Planning theory. Oxford: Pergamon.

Floyd, M. (1978) 'Structure plan monitoring: looking to the future',
Town Planning Review, 49, 476-485

Messer, I.M. & Wilson, T.D. (1983) Approaches to information management
in County Planning Authorities in England and Wales: a survey.
Sheffield: University of Sheffield, Department of Town and Regional
Planning. (Occasional paper no. TRP 40)

Van Meyel, A. (1979) 'Approaches to strategic monitoring: a case study
approach', Geoforum, 10, 387-405

Wildavsky, A. (1972) 'The self-evaluating organization', Public
Administration Review, 32, 509-520

D.R.STREATFIELD

The Education Management Information Exchange (EMIE):
a user-based research project aimed at local authority
education administrators.

"The fact would seem to be, if in my situation one may speak of facts, not only that I shall have to speak of things of which I cannot speak, but also which is even more interesting, but also that I, which is if possible even more interesting, that I shall have to, I forget, no matter."

Samuel Beckett The unnameable

INTRODUCTION

State education in England Wales, below the University level, is administered through 104 local education authorities (LEAs), or 108 if various offshore islands are included. Education is the largest government service, accounting for up to 60% of local government budgets, but the information requirements of local education administrators had received little attention before the EMIE Project was established, in January 1981.

The Education Management Information Exchange was set up as a four-year experimental project backed by the central government Department of Education and Science, the National Foundation for Educational Research, the Society of Education Officers and, more recently, the Welsh Office. Two factors which influenced this

move should be noted here. Early experience with the EURYDICE education policy information network within the European Communities (see Davies, 1979) made it evident that the decentralized British system of educational administration imposed limits upon effective international policy information transfer. Concern was also being expressed about the difficulty for education officers in learning about the experience and ideas of their colleagues in other parts of the country. A small pilot project was accordingly set up at the NFER, which houses the British 'node' of the EURYDICE network, and this demonstrated that some form of document-based information service was possible.

As a result of this experiment the EMIE Project was launched with the main aim of helping LEA staff in England and Wales to exchange information about education policy issues and useful practice. EMIE is a relatively small project, with two full-time research staff, a secretary and part-time clerical help. The staff are supported by a Principal Adviser and five part-time paid consultants, as well as by the resources of the NFER including its library and computer services. The project budget amounts to £192,000 over four years. Although this level of expenditure may be considered modest if EMIE is viewed simply as an information service, looked at in its proper context as a user-based research project it represents a significant investment in terms of British information research.

This paper will concentrate on the EMIE research activities and the strategy of involving service users in service development.

An account of the social science research issues which have affected the evolution of the EMIE Project will be published in Social Science Information Studies later this year (Streatfield, forthcoming) but reference will be made here to some of the implications of the project programme for future information work. The basic EMIE services have been described elsewhere for various education audiences (Streatfield 1983A, 1983B).

THE RESEARCH PROGRAMME

An earlier speaker referred to the field of user studies experiencing "a profound methodological crisis in the apparent opposition of quantitative and qualitative methods." (Exon, 1983) This 'crisis' might be better characterised as a profound indifference to the problems of gaining understanding of information-related behaviour through research. Most information research appears to be focussed on the information service delivery system rather than on the information user. Indeed, the concept of the information user tends to be defined by information researchers in terms of the particular information service under review, rather than regarding the information user as someone who happens to come into contact with a particular service and who may choose to use that service as one amongst many information services. To the extent that the apparent opposition of quantitative and qualitative methods has been considered by exponents of user studies, this has tended to result in a partial reenactment of the great sociological debate of the past twenty years resulting in arguments in favour of phenomenological

strategies at the expense of traditional positivist approaches. As Bell and Newby (1977) have pointed out, using phenomenological strategies to undermine traditional research methods may result, not in employment of new qualitative methods, but in having no basis for measurement at all.

The EMIE Project does not represent a case study in qualitative research. It does, however, encompass an approach to seeking greater sensitivity in interacting with system users at the service level. This involves making an attempt to illuminate a service which was conceived within the positivist tradition. At a research level, the project is based upon a commitment to learning more about how the education administrators use information in their work and about the factors that influence their use of information. It is convenient to draw a distinction between the research and service aspects of the project but, as will be shown later, these activities are closely entwined and mutually dependent.

One of several unusual features of the EMIE service is that it was only launched nationally after 15 months of preparation. This preparation involved a preliminary exploration of the field and a planning phase followed by a six month trial during which various services were tested in 30 LEAs. An important element in the preliminary exploration was a period of observation of education administrators at work.

OBSERVATION OF EDUCATION ADMINISTRATORS

Direct observation of the use made of information by people at work has seldom been employed as a research method by information researchers. This is perhaps surprising because, although it is a comparatively expensive way of conducting research, observation can be relatively unobtrusive, provides the possibility of obtaining greater understanding of the problem experienced by information users, and is a fertile source of ideas about information service development.

Three week-long periods of observation were carried out in two LEAs early in the life of the EMIE Project. The aims were to learn something about the working methods and pressures, as well as the work content, of education officers; to ensure that the relationship between information and the context in which it is used by education officers was kept in focus by the project team; and to gain some preliminary insights into communication behaviour in LEAs which would help to make the design of EMIE services appropriate.

Since the EMIE Project Head was previously employed as Principal Investigator on the major investigation of information needs and services in English local authority social services departments known as Project INISS (see Streatfield and Wilson, 1980) the choice of 'structured observation', used in Project INISS, was fairly automatic. Use of a similar approach to observation and an identical recording method enabled tentative comparisons to be

made with the Project INISS findings (which were based on week-long observation of 22 staff in five social services departments). A detailed account of the methods, experience and results of the Project INISS observation phase has been published (Wilson and Streatfield, 1980) and the observation method is described fully elsewhere (Wilson and Streatfield, 1981). It is probably sufficient for the moment to note that specially prepared cards were used to record 'information events'. An information event was defined as a subject pursued over a period of time whether through reading, writing, conversation or any combination of these. A telephone call, for example, might consist of one or more information events, depending upon how many topics were raised during the call.

MAIN CHARACTERISTICS OF INFORMATION-RELATED BEHAVIOUR

Although the three people observed were doing different jobs and were at various levels of seniority, they (and their colleagues with whom they were in contact during observation) shared a number of characteristics of interest for the design of information services:

- a) their work was fragmented with frequent interruptions, changes of topic and different participants in their communications;
- b) they disposed of most documents and publications received in a very short time, particularly when their

contact with this written material only involved reading (as opposed, for example, to referring to a document in a discussion or writing a report based on documents);

- c) most of the information events were conducted through oral (face-to-face or telephone) channels;
- d) all the people spent substantial amounts of time in previously scheduled meetings;
- e) they showed no tendency to engage in systematic and protracted searching for information on matters of current concern, but;
- f) they evinced interest in the approaches other LEAs had taken to their own current problems;
- g) they tended to seek information or advice by contacting colleagues in their own or other departments, usually through oral channels; and
- h) they showed readiness in responding to requests for information or advice in this way; but
- i) they made little use of formal information stores apart from department files.

In all these characteristics the education officers were similar

in their behaviour to the social services department managers observed by Project INISS. Some differences were, however, detected. The work of the education officers was even more fragmented than the average for social services staff (74% of information events were over in three minutes or less, compared with 62% for social services staff) and the average time spent by the education officers in reading documents and publications (84% took one minute or less) was shorter than for the social services managers (72%). Three more differences were noted:

- j) the span of responsibility held by the LEA staff tended to be narrower than that of corresponding levels of staff in social services departments, reflecting the larger size and greater complexity of education department work;
- k) the process of directly "feeding in to the policy making pool" appeared to involve a larger proportion of LEA staff, at least intermittently;
- l) the education officers observed (and most of their colleagues) were essentially programme administrators concerned with operational decisions over one or more aspects of the education service. By contrast, most of the social services staff observed were primarily concerned either with direct service delivery to clients or with day-to-day personnel management. Equivalent activities in LEAs were usually managed within the

schools and colleges rather than at headquarters level).

Using these findings, together with a great deal of supporting data it was possible to fashion a series of basic EMIE services aimed at strengthening and reinforcing the preferred methods already used by education officers to obtain information. For example, a small but carefully selected collection of LEA documents has been assembled by EMIE and details are maintained of useful LEA contacts and activities. Again, in designing an abstract bulletin service based upon LEA documents received, attention was paid to the fact that no less than 97% of 191 information events which only involved reading, were completed in three minutes or less. The result was a bulletin designed to be scanned within that time.

BUILDING UP A RESEARCH PICTURE

The observation work was not seen as an isolated activity but as part of a continuous process of knowledge building and refinement. Opportunities have been taken throughout the life of the project to talk to education officers informally or in meetings about aspects of information seeking, storage and use in LEAs and some meetings have been arranged specifically for that purpose. Invitations to observe various meetings of education officers were also accepted. At the end of each observation period a narrative report on the week was prepared for the observation subject and this formed the basis of a post-observation interview, giving an opportunity for the subject to correct errors of

interpretation or fact, comment on the experience of being observed and respond to any conclusions drawn in the report. Comments of a few senior education officers were sought to establish whether a recognisable picture of the work of education officers was emerging, and material drawn from the observation narratives was later used in information workshops for education officers, giving another opportunity to assess whether the sorts of pressures, problems and priorities encountered during observation were recognisable to other education officers.

Once the basic EMIE information services were launched (from October 1981) the emphasis in the research programme changed towards seeking interaction with potential users of the system, testing experimental services, and increasingly, evaluating the services.

INVOLVING SERVICE USERS

When the EMIE service was launched on a trial basis in October 1982, the 30 LEAs involved were each asked to nominate an education officer as a Link-person to work with the project. Meetings were held with these Link-people, who were requested to identify problems and issues which were likely to dominate their work and that of their colleagues over the six month trial period and to suggest an appropriate role for EMIE in response to those issues. Again when the project was launched nationally in March 1982, Link-people were sought in all the LEAs involved. A series of meetings with Link-people was held last summer to review progress,

and to seek views on priorities for the next year and on the services which should be offered in response. (These assessments have proved to be remarkably accurate, in terms of the demands made upon EMIE enquiry and document supply services). Three EMIE Open Days will be held for Link-people in September 1983, at which the cycle will be renewed.

A similar dialogue is maintained less formally through day-to-day contacts with education officers. Education officers are both the main users of and contributors to the EMIE service. Link-people encourage colleagues to use the services, send in documents which they consider to be potentially useful to colleagues elsewhere, and they may also be asked for information by EMIE staff when dealing with enquiries. We hope that this relationship will develop beyond the normal client-service interaction so that the Link-people will increasingly influence the development of EMIE services. EMIE now has Link-people in 102 LEAs.

THE TRAINING WORKSHOP AS A RESEARCH FORUM

Three one-day information workshops for education officers have been organised by EMIE and another three are planned. As predicted in the paper presented at IRFIS 4 (Wilson and Streatfield, 1982) these workshops have come to form an important part of the overall process by:

enabling practitioners to focus their attention on

communication issues which they might otherwise take for granted;

- allowing the project team to come into closer contact with the practical issues facing practitioners in their day-to-day work; and by
- encouraging a more equal exchange of problems and ideas between practitioners and information specialists than could be achieved by an 'information officer as expert on information provision' approach.

It was also predicted that because only a limited amount of observation and interviewing was possible in the early stages of the project the 'outsiders picture' of education work, to be offered to participants in workshops by EMIE staff, could only be tentative. This, too, proved to be correct and some initial difficulties were encountered in establishing a basis for confident communication. Extensive experience of running training workshops for social services managers had not adequately prepared the organisers for the sort of 'easy generalisation' that emerged from the small working groups of participants at the first workshop. However, recognition of the intense task orientation implicit in the 'programme administration' role referred to above, enabled more rapid progress towards a constructive dialogue in later workshops.

The most recent workshop, which grew out of the first one, was

devoted to the problems of managing information technology in LEAs and adopted a case study approach. This was organised as a collaborative venture between EMIE staff and a group of education officers.

DOCUMENT CONTROL

Two features of the EMIE service which are now sufficiently well established to have lost the initial air of experiment should be mentioned briefly. An important factor in the success of the EMIE service is the extent to which 'sensitive' documents can be obtained from LEAs and other sources. Much of this material is confidential, usually because documents have not completed their journey through the committee and consultation system. Since it is in the interest of LEA users of EMIE to have access to documents as early as possible in the committee cycle, a few documents are supplied by LEAs for use only by EMIE staff and consultants and rather more are available through EMIE only to education officers in LEAs. This approach is essential for the main users of the service but may be frustrating for other potential users interested in education policy issues. In practice, an equally significant constraint upon access to the EMIE service is that the small size of the EMIE staff allows little time for dealing with outside enquiries.

The other 'well established novelty' in the EMIE service is the attempt to introduce an element of quality control over the information being handled by employing five paid part-time

consultants drawn from the upper echelons of the education service. An important aspect of the project has been to learn how to make effective use of consultants in dealing with difficult enquiries, commenting on the content of documents, and providing overviews on topics of current importance.

EXPERIMENTAL SERVICES

EMIE has now reached a stage in its development when the success of the basic services being offered is limiting the capacity of the project team to innovate. A logical response might have been to close down the basic enquiry and document supply services since these have been shown to be successful and the project was conceived as experimental in nature. At this point the competing demands of the research project and of service development became acute since to close down the successful services would destroy the credibility of EMIE in the eyes of LEA users, but to increase the staffing level (the logical alternative, to free research staff for innovatory work) has been ruled out because such a staff increase would be seen by the main project sponsors as changing the nature of the experiment.

Attempts are being made to innovate along lines suggested by the observation findings. Recognising that the Society of Education Officers (the professional body of local education administrators) provides an important communication network through its national and regional committee activities, EMIE is beginning to provide direct information support to these committees. For example, at

regional level, a digest of information on the training of school heads and senior teachers was prepared recently as background to a regional meeting. At national level EMIE is heavily involved with the SEO Industry Committee, both by providing information on specific topics and by editing/publishing the SEO Industry Committee newsletter.

One example of the scope that exists for innovatory activity within an already established network is offered by EMIE interest in the use of computers and other forms of information technology for administrative purposes in LEAs. Having established that many enquirers were asking about aspects of this topic and that no overall picture of the state of development in this field was available, EMIE agreed to collaborate with the SEO Industry Committee in conducting a national questionnaire survey on the subject. Replies have now been received from 93 of the 108 authorities approached and a report on the survey is being prepared by EMIE. The innovatory aspect of this initiative lies in the proposal to seek official endorsement for the report from the SEO Industry Committee before passing the report to the SEO regional committees and arranging for a spokesman from an authority offering 'good practice' in each region to lead a discussion on the main findings. It is hoped that the main points raised in these discussions will be passed through to EMIE to enrich the process further and that some authorities will be encouraged to take action as a result of the discussions.

An interesting point of contrast between education and social

services departments in England and Wales is that, as a result of central government legislation to reorganise the social services, many social services departments appointed research officers a few years ago. This move was of particular interest to information practitioners because the new research officers introduced a number of changes aimed at improving the availability of information for managerial purposes (often including the establishment of small department libraries). They also organised themselves into a self-help body known as the Social Services Research Group.

LEAs have not usually created specialist research posts but a number of such posts are beginning to emerge and various education officers have some responsibility for research, planning or development work as part of their role. Following recent discussions a new LEA Research Group is now being formed by interested education officers and EMIE is offering information support to the Group.

One further example of EMIE innovation is offered in relation to new government legislation on 'special education'. Two new research projects will soon be examining aspects of local government educational provision for the disabled and handicapped and EMIE has been asked to act in a coordinating role, gathering information from LEAs on behalf of both projects as well as for the EMIE service users. EMIE is able to take on this role because the effects of the new legislation is providing a focal point for a substantial proportion of current EMIE enquiries.

In all these activities EMIE is trying to heighten the general awareness of the significance of information in people's working lives and at the same time to learn more about the problems and difficulties involved in making more effective use of information.

EVALUATION ACTIVITIES

Various elements of evaluation were built into the EMIE Project at the outset. Detailed records are kept of enquiries and how they are dealt with; recipients of information from EMIE are asked to return a response postcard commenting on the usefulness of the material received; course assessment questionnaire are distributed to workshop participants; and participants at the meetings of Link-people referred to above were asked to give their comments and those of their colleagues on various aspects of the service.

A graduate student from the Department of Information Studies at the University of Sheffield recently conducted a series of telephone interviews with 15 users of the documents supply service, treating each example of use as a 'mini case study'; similar interviews with users of the enquiry service are planned for later this year. This sort of interview, in which questions are asked about the context of the enquiry, about other services used and about the results of the search, provide a wealth of detailed information which should eventually lead to more effective service provision.

A group of link-people will be 'commissioned' to write an independent report on the EMIE service later this year to reflect their own involvement with the project and the views of their colleagues. Next year all the SEO Regions will be approached for their views on the EMIE service and other evaluation activities will no doubt be undertaken.

CONCLUSIONS

If the EMIE service is to succeed, ways must continue to be found to encourage education officers to make more systematic use of information. Even more importantly, great care must be taken to ensure that the services offered fit into the normal work pattern of the recipients. To this end, efforts are being made to gain greater understanding of how education officers use information in their work and of the factors which influence their use of information.

Some problems in developing the EMIE research programme have already been described. In an ideal world it may have been preferable to have spent more time at the outset on investigating the information world of the education officer; options other than a national information centre might then have been explored; the problem of how to balance demands created by the initial success of services in an innovatory project will need attention before the EMIE model is adopted elsewhere; and we don't yet know what effects the recent innovations are having. The project design has not allowed attention to be paid to the important questions

surrounding the use of information for strategic as opposed to operational decision making. Over the next few months the potential conflict in the EMIE staff roles as both service providers and evaluators may come to the fore, meanwhile we do not know whether the service will continue and whether other information research activities will emerge in its wake.

Whether or not the EMIE service continues there appears to be at least one important lesson to be derived from the project already. It is possible to redefine the sterile relationship between the information expert as omnipotent information provider on the one hand and the information user as passive (but not necessarily delighted) recipient on the other. By paying close attention to the normal work behaviour of potential users of an information service, it should be possible for the information provider to begin to fully live up to that name.

REFERENCES

- BELL, C. and NEWBY, H. (eds) (1977) Doing sociological research London: Allen and Unwin (See page 21).
- DAVIES, J. (1979) The Commission of European Communities' education policy information network: prospects and problems in the UK. Aslib Proceedings 31, 342-400.
- EXON, F.C.A. (1983) Methodological problems in user studies, and the organisation of in-house library research. Abstract of a paper presented at the Fifth International Research Forum in Information Science (IRFIS 5), Heidelberg, Federal Republic of Germany.
- STREATFIELD, D.R. (1983A) Information service for education officers Educational Research News 38 (Spring) 3.
- STREATFIELD, D.R. (1983B) Meeting demands for information Education 161 (May 27th) 412.
- STREATFIELD, D.R. (forthcoming) Moving towards the information user: some research and its implications Social Science Information Studies.
- STREATFIELD, D.R. and WILSON, T.D. (1980) The vital link: information in social services departments London: Community Care and the Joint Unit for Social Services Research.

WILSON, T.D. and STREATFIELD, D.R. (1980) You can observe a lot ... a study of information use in local authority social services departments Sheffield: Department of Information Studies, University of Sheffield.

WILSON, T.D. and STREATFIELD, D.R. (1981) Structured observation in the investigation of information needs Social Science Information Studies 1, 173-84.

WILSON, T.D. and STREATFIELD, D.R. (1982) Action research and users' needs in Friberg, I (ed) The 4th international research forum in information science: proceedings Boras, Sweden: Hogskolan i Boras.

A DATABASE MODEL AS THE BASIS OF A WORKSTATION FOR THE DESIGNER

R. McCall, Jr., B. Schaab, W. Schuler and I. Mistrik

Gesellschaft fuer Information und Dokumentation mbH
Sektion fuer Systementwicklung
Heidelberg, F.R.G.

1. INTRODUCTION

In recent years it has become clear that microcomputer-based workstations will soon play an important role in supporting the work of various types of professionals. Many researchers are now at work developing such systems as aids to all sorts of human tasks. We are currently engaged in a project to develop a microcomputer-based workstation for a type of problem solver we call the designer. In this paper we develop general set of requirements for such a system together with the design-theoretical considerations from which these requirements are derived.

Our theoretical starting point is Horst Rittel's theory of design. Our work here is intended as a compatible extension of that theory and of the IBIS (Issue-Based Information System) method, which derives from that theory. Briefly summarized, our argument is as follows. Rittel's theory points out that design problems are characterized by openness and controversy. Because of this the designer finds himself in a dilemma for which his only solution is a "radical" purposefulness. As a consequence the designer needs an information system, but one which is fundamentally different from conventional information systems because it must handle a fundamentally different sort of database, a database characterized by having a quasi-hierarchical functional structure, queries within the database itself, and a variety of information flows into, within, and out of the system.

2. THE NATURE OF DESIGN

It is appropriate that we begin by defining the terms "design" and "designer." By "design" we mean the activity of aiming at the production of a plan which is intended if implemented to achieve certain goals without creating an unacceptable constellation of side- and after- effects. This definition, a slight variation of Rittel's, is an extremely broad one. Among "designers" we include people in a wide range of fields, such as architecture, city planning, product design, information system design, and software engineering. Researchers and planners working on development projects whose outcome has an impact on society or on the environment are also included. It has often been supposed that design problems are like the problems of mathematics or science and therefore could be usefully approached in a mathematical or scientific manner. Rittel, however, has argued that design problems are fundamentally different from these well-defined but often complex problems. As he sees it design problems are not merely highly complex, they are fundamentally ill-defined, ill-structured and ill-behaved. In his terminology design problems are "Wicked Problems" (WPs) [2].

Rittel lists ten properties of Wicked Problems, but for the sake of simplifying the present discussion we can say that WPs have

two main characteristics of which most of the other characteristics identified by Rittel are aspects. These two characteristics we call "openness" and "controversiality" [6].

Openness

A WP is open in the sense that the set of data understood to be relevant to the problem remains open to new input throughout the design process. There are four reasons for this: 1) the formulation of a WP is inherently open-ended; 2) there are no "classes" of WPs; 3) the opportunity for human judgment and creative insight exists throughout the design process; and 4) much design information comes from people other than the designer.

Open-endedness: The formulation of the problem itself (and therefore its solution) continually develops during the design process and could (in theory) go on forever - though of course some cut off point must be found.

No "classes" of WPs: Wicked Problems are not really separable from their environments. Though there can appear to be important commonalities between design problems, it is difficult to say that the same design problem ever really occurs twice, for the particular constellation of factors that affect problems may never be repeated. And this constellation can make a project different from apparently similar projects in overridingly important ways. For this reason it is not possible to develop algorithms in advance for solving certain types, or classes of design problems.

Openness to judgment and creativity: The design effort is open to the input of human judgment and creative insight throughout the design process. Finding the best solution for a WP is not merely a process of "decision making", i.e., choosing from a set of given means. The set of means need not be restricted to those already defined, but can be and often should be found or invented by the designer. The designer can use his creative insights to reformulate the problem completely - if, for example, the solutions to the originally formulated problem do not seem adequate.

Openness to information from others: The designer has no logical or ethical basis for cutting himself off from the rest of humanity when he works. In fact, acceptance of the definition of design proposed here, i.e. the production of a plan which does not create unacceptable side- and after-effects, implies that he is both logically and morally obliged not to isolate himself. This is in contrast to the situation of someone working on a purely mathematical problem, where isolation is logically justified and may be desirable. Unlike the goals of a purely mathematical problem, the goals of a WP are not abstractions. They are real effects on real people. Not only does the designer have clients, he must consider the effects of his work on other people as well. Ultimately his task is to find out who is affected and how. In order to do this he must have contact with

people who might potentially be affected and try to elicit information from them.

Controversiality

WPs are inherently controversial in the sense that formulating and solving a WP cannot be done by procedures which are objective or in any sense automatically beyond dispute. There are two main reasons for this. The first is that design involves deontic information - i.e., information about what is good or bad, appropriate or inappropriate, what ought or ought not to be. Every step of the problem formulation requires such information. Deontic information is notoriously the subject of disputes for which there are no objective arbiters. The second reason is that factual information used in design often fails to meet the criteria for scientific testability and therefore remains in principle disputable.

The immediate implications of the openness and controversiality of WPs are as follows. The designer is responsible for the larger consequences of his acts: for the side- and after- effects and also for the legitimacy of the goals he attempts to accomplish - legitimacy both in the sense of the higher goals of his client and the welfare of the larger systems in which the clients' problems are imbedded - ultimately the commonweal. To accomplish his task the designer needs to get opinions representative of the persons affected by the outcome of the project. In addition he needs a wide range of opinion on all matters he deals with. This requires that he take dispute into account.

The Design Dilemma

Because of the wickedness of design problems the designer is a person who is by very definition in a dilemma. He has two important but conflicting aims. On the one hand, he wants to obtain, record, and further develop the best information about what the plan should be. He could, at least in principle, go on indefinitely trying to document and develop more and better information. On the other hand, he needs to complete his project and do so within the limits of the resources at his disposal. Failure to do this would render worthless all of his efforts to improve his planning by obtaining better information. Dealing with this dilemma is the central methodological difficulty of design.

A Radical Teleology

The problem with which the design dilemma presents him is that of determining how best to allocate inadequate resources to an important task. The good designer's reaction to this problem is to become deeply concerned about maximizing the efficiency of the use of his design resources - i.e., of maximizing the return - in terms of design quality - of his investment of time, money, manpower, intelligence, etc. in the design task. He tries to use

his resources in the most productive way - where productivity is measured in terms of the quality of the information produced, and the quality of the information is measured in terms of the quality of the plan the design process results in.

This deep concern for efficient use of design resources thus leads the designer to develop a radical attitude toward design procedures and design data. He becomes highly intolerant of things which might waste his resources or interfere in any way with his obtaining the information he needs. He is a merciless in winnowing and pitiless in making priorities. At any given point in a design project, he is reluctant to devote any resources to the collection or processing of data whose importance to the project is not demonstrably great. Yet he is unwilling to forego the possibility of having access to such data later in the project.

We call this a "radical" commitment to teleology because it is extreme, single-minded, and uncompromising. It puts no faith in methods, models, theories, or information unless they are of demonstrable value to the design project at hand. The designer uses such things whenever they will help him. He unsentimentally abandons them if they fail to be helpful for his immediate purposes.

As we show below, this radical teleology is the single most important factor in determining what the design of a workstation for the designer should be.

3. THE DESIGNER'S NEED FOR AN INFORMATION SYSTEM

As we have seen, the openness and inherent controversiality of Wicked Problems place the designer in the dilemma of having to gather, organize, evaluate, and use large amounts of information - including conflicting information - with resources which are inherently inadequate for the task. The good designer's reaction to this situation is to go about this process with a "radical" efficiency.

The good designer's goal at every stage of the design project is to obtain quickly all the information but only the information which is relevant to the particular design subtask with which is currently concerned. Yet the collection of data with which he must cope is so large and the tasks so varied and changing that he cannot satisfy this goal without support in the form of a computer-based information system.

The designer needs an information system, but more than a conventional information system - for conventional systems are closed and avoid controversy. The designer needs an open information system to deal with his open problems. The system must be open to the input and generation of new data throughout the design process. It should help to obtain as well as to contain information. It should allow and encourage input of ideas from the designer, those affected by the outcome of the design

process, and others qualified to participate. Conventional information systems avoid the controversial, but the designer's system must embrace it. It must allow disputable data. It must promote communication between the designer, his clients, and others who may be involved in the design process. The system must be able to manage, structure, and sustain the discussion among these people. The information system must help the designer to decide how to allocate resources. It must help him decide what tasks to undertake and how much in the way of resources to devote to each.

4. SPECIAL CHARACTERISTICS OF DESIGN INFORMATION

By design information we mean the data which is informative with respect to some design project. In other words data which can inform the designer in his task of devising a plan which if implemented will achieve certain worthwhile goals without creating an unacceptable constellation of side- and after-effects.

According to the definition we have given of design, the set of design information could include data about the goals to be achieved, the possible side- and after-effects of means which might be employed to satisfy the goals, and the various means which might be employed. But this is not all. Any data used to devise, find, or evaluate design information could also be design information. Thus any thought or discussion about the project could become design information.

Design information is fundamentally different from the sort of information found in conventional databases. In particular, it has two special characteristics which must be taken into account by an information system for designers. One of these is that design information is only information with respect to the individual projects in which it is used and with respect to those project-subtasks for which it is useful.

When these tasks are completed, the data which was relevant to them can no longer be informative for these tasks. Thus the ability of design data to inform the designer with respect to a given task is destroyed by the designer's use of that data for that task. In other words, design information is consumed. At the end of a design project none of the data used is design information with respect to the project, since it is completed.

The second special characteristic of design information is that it is often generated during the design process itself, and not merely obtained from existing collections. This is in part because the tasks performed by designers, being 'wicked' in nature, require the development and modification of ideas by the designer himself. In addition to this he must extract information from a variety of sources available to him in order to meet needs which are as unpredictable as his problems are unprecedented. He has to be a critical evaluator and judger of all available information in his environment or offered from

elsewhere. Much information first arises during discussion and communication among various contributors of information to the project.

The designer's system must allow and support both these characteristics of design information, i.e. the fact that with respect to a given task design information is consumed and is thus only temporarily informative, and the fact that it is actually generated and developed during the design process itself, which means there is a constant influx of potentially relevant data into the system.

5. A MODEL OF THE DESIGN DATABASE

We will develop a model of the designer's database which makes no assumptions about how this database is to be implemented. From this model we will derive certain conclusions about the requirements for a design information system. There are two different perspectives from which to view the designer's database: 1) functional structure and 2) use status flow. We will first discuss these separately and then synthesize them into an overall database model.

Functional Structure

From the "wickedness" of design problems we can deduce that design information has a characteristic distinguishing it from information represented in conventional databases, which is that it has a functional structure. By this we mean that there exists a set of relationships between information items which indicates the function of the information, i.e., how the information nets used in the design project. Our argument is as follows.

The designer's radical teleology requires highly efficient storage, retrieval, and use of information. He wants to be confronted with as much data but only as much as he can process with the resources at his disposal. But this data must be both relevant and of the highest priority.

What should the design of his information system be in order to help him to meet these goals? The key to answering this question is to be found in certain ideas on which the IBIS and PHIBIS design methods are based. Although any such method is prescriptive, the ideas we are referring to are descriptive. To be specific, they are insights about inevitable structural features of design information. In order to better explain these features we will briefly explain IBIS and PHIBIS.

IBIS (Issue-Based Information System) is a method for designing by individuals or groups [1] which breaks the overall problem-solving process down into subprocesses consisting of the raising and answering of various questions, which are called issues. Answering these issues involves 3 basic steps: 1) proposing various answers, 2) arguing the pros and cons of these proposed answers, and 3) determining, on the basis of the

argumentation, which proposed answers are to be accepted or rejected. The documentation of the issues, proposed answers, argumentation and decisions is central to the method and is used to coordinate and to stimulate discussion.

The data handled and processed in an IBIS consist mainly of the actual statements (texts) of issues, positions, and arguments. These data are related to other data necessary for system monitoring, etc. such as authors, dates, keywords, and references to literature or experts as well as information on the status of single issues, i.e. who is responsible for resolution, the deadline for resolution, or to which extent the issue has been treated, etc.

IBIS is based on the important descriptive insight that design can be considered a question-answering activity. Questions - i.e., issues - are thus the central units of design information and all other design data issue-related. In other words, the way in which the data is informative is in its role in helping to resolve some design issue. IBIS also provides categories of issue-relevant data: answers, arguments, authors, and so on.

PHIBIS (Procedurally Hierarchical IBIS) combines the IBIS approach with the concept of procedural hierarchy and the associated concept of stepwise refinement - both of which have become well-known by their association with computer programming [7]. The concept of procedural hierarchy is implemented in PHIBIS by means of a new principle for relating and developing issues [3]. Basically, the principle is as follows. First an overall issue for a given project is developed. For example, if the project is to design a database management system, the overall issue might be phrased, "What should the design of this database management system be?" Just as a problem can be broken down into subproblems, this first issue is broken down into subissues. Examples of such subissues might be "What sort of information is to be handled by the system?" and "What sorts of displays of information are needed by the users?" and "How frequently does data need to be updated?" Each of these issues might then in turn be broken down into subissues - i.e., subissues of the subissues. This process of breaking issues down into subissues continues until issues are developed which are answered directly - i.e. without the need to raise subissues of these issues.

This process results in a tree-like structure of issues in which the process of resolving the subissues of any given issue is part of the process of resolving the given issue. In other words, the processes of resolving the various issues are nested in a procedural hierarchy, in which ultimately all issue-resolving processes are part of the process of resolving the first issue: the overall issue. The inter-issue relationship in such a procedural hierarchy we call the serve relationship. It is defined as follows: Issue A serves Issue B if and only if resolving A is useful for resolving B. Or put another way: the answers to Issue B depend on the answers to Issue A. (The resolution of Issue A contributes to resolving Issue B).

The central PHIBIS insight is that each design project can as-a-whole be thought of as the process of resolving a single "overall" issue. In other words, for any given project it is always possible to formulate an issue such that the resolution of this issue IS the whole project. This means that all issues besides this "overall" issue are by definition resolved for the purpose of resolving the "overall" issue. In fact, it is only by functioning in this way that a question can become an issue in the design project. Thus PHIBIS takes the IBIS logic one step further by showing that issues themselves are issue-relevant data. Thus all design information, with the exception of the "overall" issue, functions as information by being informative with respect to the answering of some issue. This is furthermore the only way in which data can be design information.

A database of design information thus has the important characteristic of consisting in part of the questions which a designer wishes to answer. This situation is to be contrasted to the situation with a conventional database, in which the questions that the user wishes to answer - i.e., the queries - are external to the database. The purposes for which design data are used are themselves represented by design data. Thus an important characteristic which differentiates a design database from other sorts of databases is that almost all the data in the database has an informative function with respect to other data in the database. In other words, in a design database there is a network of interconnections between data item which corresponds to the functional structure of the database - i.e., a structure which indicates how the data is informative by showing what data it is informative with respect to. The way in which each data item is linked in the structure makes it clear how this item contributes to the designer's purpose. Arguments serve to evaluate answers in the resolution of issues; the resolution of a given issue aids in resolving other issues, and ultimately the "overall" issue. This functional structure is fundamentally different from the semantic or formal structures of conventional databases.

A second insight from PHIBIS is that the functional structure of issues in a design project is quasi-hierarchical - i.e., more or less tree-like in structure. It must be stressed here that the PHIBIS method does not prescribe a quasi-hierarchical structure. The existence of such structures in design is a descriptive claim based on empirical evidence. PHIBIS further advocates the top-down/stepwise refinement approach to the development of the issue structure, an approach also currently dominant in software engineering (5,6).

Figure 1 shows the sort of functional structure - i.e., serving structure - that exists among issues, while Figure 2 shows the functional structure relating non-issue data types to issues.

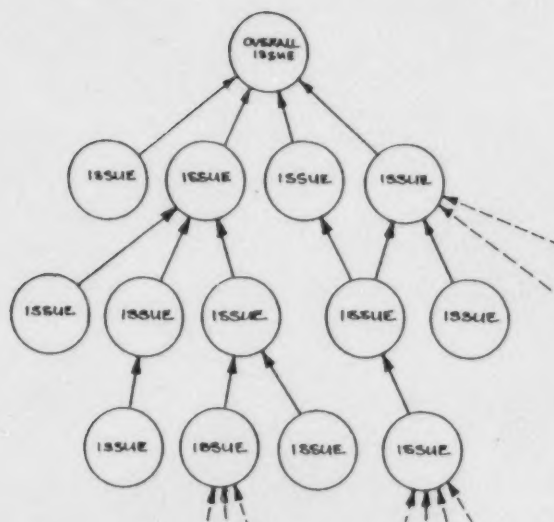


Figure 1. The Functional Structure of Issues.

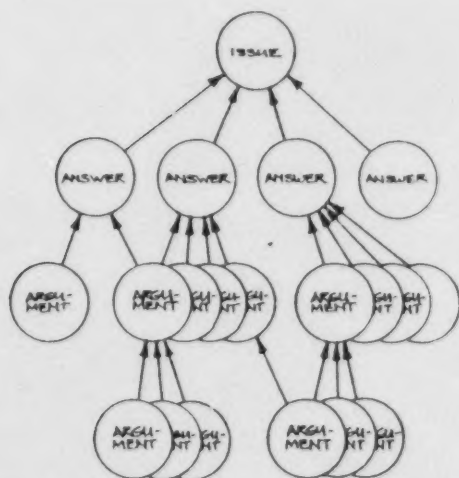


Figure 2. The Functional Structure Relating an Issue to Other Types of Data.

Use Status Flows

A second way of categorizing data items in the system is by what we call its use status. As shown above, all items of design information, with the exception of the "overall" issue, function - i.e., are used - as information with respect to some other piece of design information. Conversely, for any given item of design information there might be a number of other items which are of conceivable use with respect to it. These items can be in the following mutually exclusive use-status categories:

1. ACTIVE:
 1. PROPOSED: proposed for possible use but not yet evaluated to decide whether it is to be used.
 2. UNDER EVALUATION: under evaluation to decide whether it is to be used.
 3. ACCEPTED: accepted for use but not yet in use.
 4. IN USE: being used for the evaluation or generation of other design information.
2. RETIRED:
 1. REJECTED: rejected for use.
 2. USED: no longer in use.
3. LATENT: waiting to be made into a specific proposal for use.

It will be noticed that the active status is a somewhat complicated collection of four use-status subcategories. The first three of these are untypical of non-design information but are necessitated by the open-endedness and controversiality of design problems. The controversiality means that judgment must be made on the reasonableness of data proposed for use, this often being influenced by other design information in the form of arguments. In addition, the designer's radical teleology leads him stringently to screen proposed data for relevance and priority.

It is clear from the names of the use status categories that a items of information can change status, that is, move from one category to another. Thus, as one can expect, there tends to be a general flow of information from latent to active to retired and that furthermore information flows into the latent category from outside the set of categories. Also, within the active category there is a general flow from proposed to under evaluation to rejected and/or accepted and from accepted to in use to used.

These flows (shown by the solid arrows in Figure 3) are not the only ones between use-status categories. Other flows result from the fact that all unused data which is not explicitly rejected loses its chance to be useful if the data for which it was thought to be useful has already been evaluated or used. Such unused data is implicitly rejected by being neglected. Thus, for example, a designer may refuse to consider certain arguments on a proposed answer if he thinks that he must make an immediate decision on this proposed answer. These arguments are implicitly rejected as are any arguments on these unconsidered arguments. This implicit rejection is represented by the dashed arrows in

Figure 3.) It should be noted that by the end of the design a project all data has flowed into the used or rejected categories.

Data need not pass through all the steps in the sequence indicated by the solid arrows. It can bypass certain steps and can enter the set of use-status categories at almost any point (as indicated by the dotted arrows in Figure 3).

Finally, it should be noted that despite the fact that data in the used category has been consumed with respect to the given purpose, this by no means implies that it is useless. It is quite possible that it will be useful for other purposes in the project. This is a point to which we shall return in the discussion of synthesis.

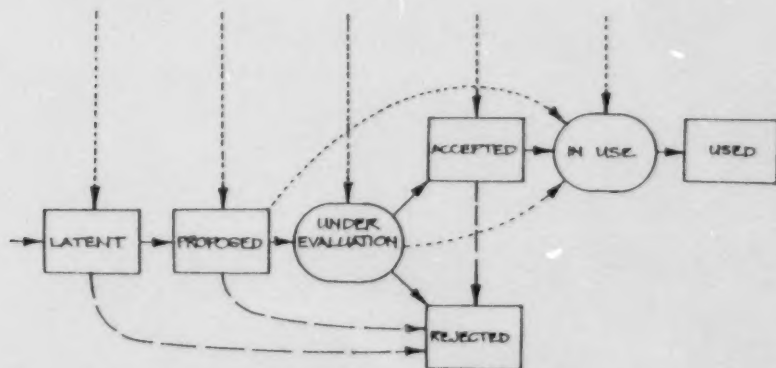


Figure 3. The Flow between Use Status Categories

The significance of the use-status categories is that they indicate what information the designer should and should not be confronted with at various points in the design process. In particular, it will be noted that two of the use-status categories correspond to processes while the rest do not. (In Figure 3 the process-related status categories are indicated by ovals.) These categories are in use and under evaluation. For each of these appropriate data displays are needed.

For the process of use of data with respect to a given other data item the designer should not be confronted with any retired data, since this is not capable of informing him. He obviously needs to see all accepted data. He also needs an indication of the existence and quantity of proposed and latent data so that he can judge the investment of resources that would be required for the

preparation and evaluation of this data.

For the process of evaluation of proposed data the designer needs a display of just that data which needs to be evaluated and an indication of the existence and quantity of latent data. For the evaluation he also needs for each item to be evaluated a display of just that data which is useful for that evaluation. Thus, for example, when evaluating a proposed answer he needs a display of all those arguments but only those arguments which are relevant to that answer.

Synthesis

Having identified the two major ways of categorizing and relating design information we can now combine them to derive an abstract model of a design database as a whole. This model will then constitute the basis for an information workstation for the designer. The model will consist of purely abstract components, by which we mean individual data items and sets of data items, together with two types of interconnections among components: 1) functional relationships and 2) information flows. The way in which this model provides a basis for an information workstation is by indicating how the workstation enable the designer to categorize and relate data, how it must allow modification of categorization, and what sorts of displays it must be capable of delivering. In other words, the model constitutes an abstract specification of requirements that a design information system must fulfill if it is to satisfy the designer's information needs.

Components of the System

Besides having individual items as components, the model needs to have a set for each way of classifying information by both function and use status. For function this means that it must be classified by both the item it functions with respect to and the role in which it functions. Thus, for example, items could function with respect to a given issue in the role of subissue. For each such functional role with respect to a given item the system must have a component - or at least potential component - for each possible use status. Thus for each issue it must have the possibility of having sets of latent, proposed, accepted, rejected, and used subissues as well as sets of subissues under evaluation and in use.

The mere combination of functional categories with use status categories does not, however, provide all required sets of data. So far we have treated use-status and function as if they could be discussed independently. In fact they cannot, because data is often latent precisely because it has not yet been assigned a specific functional role. It often happens that all that is known about a given item is that it is relevant somehow to a given other item. Therefore the system must have for each item a potential set of latent data which is relevant to it in some as-yet-unspecified way.

It should be noted that the specification of required components is not a static one. Since information flows into the database during the course of the design project, the number of components the system requires increases as the project progresses. The sizes of the individual components also change during the project, and some such change in the database occurs at each step in the design process.

Interconnections of Components

We have already described the functional interconnections of data items and the general flows between use status categories for the object useful with respect to a single other data item. But there are also flows between the use-status categories for different items. Two cases are particularly important. They are the case of 1) multifunctional information and 2) functional specification.

Multifunctional Information

As pointed out before, just because information has been used and thus consumed, i.e., used up, for a particular purpose, this does not mean that it is useless. In fact, it may very well be useful for some other purpose. In other words, it may very well be useful with respect to some other data item. In fact there are certain information items which are useful for many purposes. These include such things as principals and standard evaluation criteria. In the case of such information its appearance in the used category for one data item may cause the designer to put it in some other use-status category for other data items. Thus an argument on an answer might contain a general evaluation criterion. After accepting and/or using the argument, the designer might propose its use for the evaluation of other answers to the same issue.

Functional Specification

As mentioned above, latent data of consists of data which is relevant to a given item in some unspecified way. This relevance to a given data item is not itself an explicit functional role, but rather an expectation that a specific role can and will be assigned. Specifying what that role is might merely be a process of assigning one of the possible functional relationships - subissue, answer, explanation, etc. - to the allegedly relevant item. Very often, however, a more complex process is involved. Determining how one item is relevant to a second might involve specifying the functional role of the first with respect to a third item. This can happen when the third item itself plays some direct or indirect functional role with respect to the second. Thus for example, the designer might judge a certain statement to be relevant to a certain issue and then later determine that this relevance consisted of being an argument with respect to an answer to the issue.

The process of functional specification can be even more

indirect. In the most extreme cases the designer merely believes an item to be relevant to the project but has not yet determined how. In this case the new item is at first only indicated as begin relevant to the project's "overall" issue. Later a specific use for it must be found somewhere in the functional hierarchy.

It should be noted from these examples that the specification of the functional role of an item always results in its being transferred to a use status category of an item lower in the functional hierarchy.

A second complication in the process of functional specification is that it may not be able to be accomplished in one step. Instead of being able to immediately assign the relevant data a specific functional role with respect to some lower-level item, it might turn out that the designer can only indicate it as relevant to some lower-level item. In this case the functional specification process must continue at some future time if the allegedly relevant item is ever to get used in the project. It can turn out that functional specification is a quite complex, iterative process in which the relevant data item drifts ever lower in the functional structure, always traveling from the latent category of one item to the latent category of another until it finally lands somewhere where it can find a specific use.

When the designer comes up with any new data it may be clear from the outset which specific issue or decision the data influences; on the other hand, he might have only a general idea about which part of the project the information is important for. He may only know that it is somehow relevant for the project as a whole. In other words, data can be latent at any level of the functional structure. Later, when more specific issues have crystallized and the issue discussion has been further developed, the data can be moved to the level at which it is needed.

Flow into the System

So far we have mostly discussed flows within the database. It is, however, a important feature of design that there is flow into the database all during the project. As explained above, the criteria for entry of data into the system are that it be important for the project at hand. In other words, data must be relevant, of high priority, and not already in the information system. Note, however, that reliability in the sense of being undisputed is not necessarily a criterion.

Data should be able to enter the system from a variety of sources. One source is the designer himself. His judgments, opinions, and creative ideas are themselves representable as data of the utmost relevance and priority for the project at hand.

A second source for the designer's information is the people other than himself who are interested in the outcome of the project. The most obvious members of this group are the clients -

i.e., the people providing the funding. Sometimes those who supply the funding are not the same as those whom the system is supposed to serve. In such cases the latter constitute another important group of people. Yet another group is those who are affected by the outcome of the project even if they are not the targetted users of the system. A final group is others who might be interested in the project.

The designer can get useful information without getting it directly from people. He can get it from books, journals, magazines, or even from other information systems - though in fact very little of the information that today's designer uses is written down anywhere.

Flow out of the System

Very little of the data that gets past the latent stage should ever be discarded even when it is used or rejected. We have mentioned that used data is consumed with respect to the purposes for which it is used and that the designer should not be forced to see such data when he is concerned with these purposes. But just because data has been used does not mean it is useless; it could be useful for other purposes. There are four ways in which used data can be useful. One is by being multifunctional, i.e., useful with respect to other data in the project. A second is by being used to explain and defend decisions made. A third is by being useful for future projects; once the designer has gone to the trouble to find, edit and evaluate useful information it would be a waste to throw it away if it is of conceivable use for future projects.

The fourth reason for not discarding used data is that it might be reused for the same purpose. No design decision or design task can be guaranteed to be truly final until the entire project is complete. Thus all information which has been used in the project can be potentially be useful at a later date.

The same four reasons dictate that rejected data not be thrown out. It is conceivable, though not likely, that what is rejected for one purpose in might be useful for some other purpose in the project. Rejected proposals can be useful for explaining and justifying decisions. Proposals rejected in one project might be useful accepted in another project. And finally, the possibility of the designer's changing his mind about decisions means that rejected proposals might need to be reconsidered.

There is, however, one kind of rejected data which should generally not be kept, namely, data rejected as irrelevant. Such data should be eliminated in order to avoid overloading the database with garbage useless for the designer's objectives.

6. SUMMARY

The designer's radical teleology requires a radically different kind of information system. Unlike conventional information

systems the designer's system is characterized by

- 1) containing disputable data;
- 2) containing queries within the database;
- 3) having great amounts of information flow into the system and within the system;
- 4) having a database with a functional structure.

The conceptual structure of the conventional system in contrast is essential a static one. It behaves like a "storage tank" into which the information is poured before being used. Here information does not flow from one place to another in the system. Nor does it flow in or out during use of the system. Data in a given category remains in that category during use. Moreover, in a conventional system the user is merely a consumer of pre-prepared and prepackaged data. In the system we propose the user - i.e., the designer - is responsible for entry of data into the system. He must obtain the data and judge whether to include it. He frequently produces information himself.

7. REFERENCES:

1. Kunz, W., Rittel, H., Issues as Elements of Information Systems, Center for Planning and Development Research, University of California, Berkeley, Working Paper No. 131 (July 1970).
2. Rittel, H., Structure and Usefulness of Planning Information Systems. *Bedriftsøkonomien*, No. 8 (1972), 390-396.
3. McCall, R., Jr., Structure and Use of Issue Systems in Design, Ph.D. Thesis, Dept. of Architecture, University of California, Berkeley (University Microfilms, Ann Arbor, Michigan, 1978).
4. Bowles, K., Problem Solving Using Pascal (Springer, New York, 1979) 281ff.
5. Bowles, K., *Ibid.*, 282.
6. McCall, R., Jr., Mistrik, I., Schuler, W., Integrated Information and Communication System for Problem Solving: Basic Concepts, in: Glaeser, P. (ed.), Data for Science and Technology, Proceedings of the Seventh International CODATA Conference (Pergamon Press, Oxford, 1981), 512-516.
7. McCall, R., Jr., Schaab, B., Schuler, W., Information Station for the Problem Solver: Systems Concepts. In: Proceedings of Int. Conf. on the Application of Mini- and Microcomputers in Information, Documentation and Libraries. Tel-Aviv, Maerz 1983. North Holland Publications, Amsterdam (in press).

ONLINE MAN-MACHINE INTERACTION FACILITIES : A COGNITIVE VIEW

By Peter Ingwersen
Research Fellow
European Space Agency
Information Retrieval Service
Frascati - Italy

ABSTRACT

The paper discusses the impact of selected command language facilities on the man-system interface in relation to operational online Information Retrieval (IR).

The concept of data and information is considered in relation to the cognitive viewpoint, Brookes' and Popper's fundamental schemes.

A cognitive IR model is briefly outlined, followed by a discussion of differently experienced searchers' knowledge structures. The characteristics of online command languages are described in view of the information concept. Recently developed online search facilities such as use of contextual logic (free text operations), crossfile searching and term frequency analysis (zooming) are discussed in relation to the IR process. The cognitive view is applied as the means to describe and emphasise the retrieval possibilities and the state of knowledge involved in the interface situation. Finally, more developed man-machine interactive enhancements and their implications are suggested.

1. - INTRODUCTION

At present the global information environment appears to be in a transition period. In particular, the online industry is beginning to realise that certain basic views and assumptions about IR will have to be revised. There are several reasons for this.

Most important is the challenge posed to the industry by personal computer, videotex and word processor manufactures. Already today, inexpensive micro software packages demonstrate capabilities and flexibility equal to mainframe software. With increased memory space the microcomputer is going to definitively change online IR into a new game.

In addition, the commercial (or semi commercial) online IR systems are becoming highly complex and interdependent of the surrounding society. The demand for and production of data seems ever growing.

New user groups are evolving. To the traditional user community deriving from scientific, technological and business fields will be added the ordinary citizen with his home computer or videotex equipment. Not like a decade ago, user groups now consist of persons covering a very broad range of learning and educational levels - also concerning IR knowledge.

New technological possibilities, the increasing amount of records and the IR systems' complexity as well as the need for relevant data posed by a growing body of users to IR systems have a great impact on the man-machine interface. This interface situation is controlled by the command or search language possessed by the IR system. Two crucial factors on which the success of today's online searching depends are the searcher's ability to cope with the functions and effects of the command language facilities and his understanding of the data output.

The aim of this paper is to analyse some selected search language features and their impact on the user-system interface in online retrieval, seen from a cognitive point of view.

This view states that it is the individual state of knowledge of the information processing device - whether human or machine - that mediates the actual information processing (1). This paper discusses which kind of knowledge structures (KS) is involved and how the information processing is carried out by the participating KS in online IR using such facilities.

Within this framework it becomes necessary to distinguish between data and information. Furthermore one must emphasise that present operational IR systems consist of passive or fixed KS, embedded in mechanisms like database structures, command language syntax and indexing rules - the system's setting.

The only active and changeable KS in the interaction are the knowledge structures of the searcher. So far, no machine-man interview takes place and no mapping of need statements are carried out automatically.

The following features are related to the IR process and exemplified using the ESA-QUEST (ESA-IRS) command language:

- free text searching using contextual logic
- crossfile searching (QUESTINDEX)
- term frequency analysis (ZOOM)

The first two are common to search languages available on most operational systems. They are quite traditional system spin-off effects, designed to cope with the increasing number of accessible term relations, records and databases.

The use of contextual logic and crossfile searching are typical representatives for what Belkin et al. call the "best match" principle (2). Both imply the basic assumptions that the user is able to specify the information required - and that the information need expressions are functionally equivalent to document texts, i.e. equivalent to information.

The ZOOM facility is based on a different and more realistic view: the user is rarely able to specify the information required - and rarely in a suitable form that matches the system; at least not in one single statement. From IR studies it is known that initiating query statements often take the form of a "label" containing few terms, conceptually far from the underlying need (3).

By displaying ranked concepts from an entered set of references ZOOM may support the user restructuring or specifying the information need expressions. ZOOM allows alternative ideas to emerge and serves as a relevance feedback mechanism.

Although based within the framework of the "best match" principle, ZOOM points to more developed online trends that recognise flexibility, contextualisation, simplification and automation as vital approaches to online retrieval.

2. - DATA and INFORMATION

The central point in any information processing, according to the cognitive view, is that "the processing is mediated by a system of categories or concepts, which - for the processing device - are a model of its world" (1).

This world model is named "Image" in fig. 2 and consists of knowledge structures.

De Hay has defined "information" as seen as "supplementary to a conceptual system that represents the information processing system's world knowledge (4). The system's world knowledge - its knowledge structures - provides the basis for decisions on which ambiguities (or problems) should be eliminated.

This implies that "information" is that knowledge - or conceptual context - which is necessary to be present as a part of the processing system's KS to solve ambiguities or problems. Consequently, "information" is knowledge which has to be perceived or produced by the system. Otherwise, the actual problem cannot be solved completely. Thus, the system may try to find the information, either by thinking i.e. restructuring its KS, or by obtaining it from outside itself, i.e. to perceive knowledge from other systems.

This notion of "information" is similar to that of Brookes who, so impressively, introduced the 3-world model of Popper and its relevance to information science (5). It was also reflected by Mackay in 1960 (6).

Brookes' Equation

$$K(S) + AI = K(S + AS)$$

is very useful since it stresses what happens if a processing system (e.g. a person) perceives and interprets "something" which thus becomes information.

This "something" is data. If not perceived (or mediated) it remains data. To perceive data the processing device requires KS - category or concept structures - that are able to interpret the data. This may happen via recognition. From studies of the memory (8,9) and the learning process (10) it is known that we memorise, learn and think by the means of concepts. To supply a KS with new concepts or concept relations implies that they somehow are linked conceptually to recognised structures.

Concepts are defined by Gowin as "perceived regularities in events or objects as designated by a sign or symbol" (11).

One may therefore suggest that Popper's World 3 - the world of objective knowledge - is "potential information" consisting of beliefs, intentions, ideas, theories... in form of (objective) conceptual KS. See fig. 1a and 2.

When World 3 is accessed the "potential information" is data in the first place. The data is communicated designations, i.e. signs, symbols, words, text ... that contain what Schank calls meaning and inference (8) and the French philosopher Merleau-Ponty called the "silence" or "thought" (12).

In the event of perception, the data is transformed by the actual KS into information. Some potential information, however, does not represent new knowledge. Concepts or concept relations may, as stated above, be recognised, being previously stored in the memory. It is the new information that may transform the KS or act as a clue to modification. How the information affects the KS depends on the state of knowledge of the individual and the complexity of the information perceived (ΔI , fig. 1).

In view of this, Brookes' formular may contain in arrow and not an equation sign:

$$\Delta I + K(S) \longrightarrow K(S + \Delta S)$$

in which ΔI is recognised and new information and ΔS is the impact of ΔI on the former $K(S)$.

Fig. 1 outlines some key steps in this highly complex process. Fig. 1a shows how data is transformed into information and fig. 1b demonstrates Brookes' formular embedded in information processing.

Brookes' formular is seen as representing (human, subjective) information processing, based on supplementary data input, which is a vital part of the cognition process. One may merge Popper's scheme concerned with the process of understanding and problem solving (13) and Brookes' formular. To transform Popper's Tentative Theory State - TT - into a state of Error Elimination - EE - is in general required additional information by the subject, see fig. 1c.

Fig 1. Information retrieval, processing & problem solving

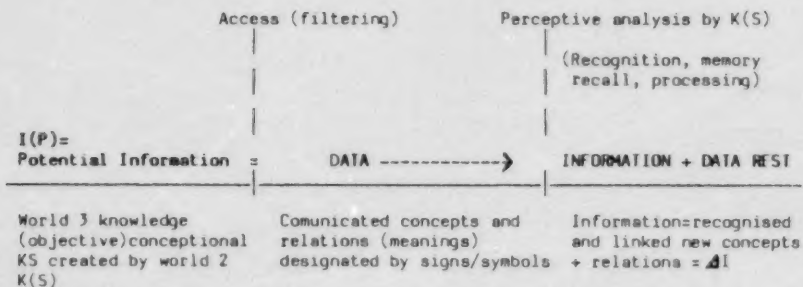


Fig 1a: Information Retrieval

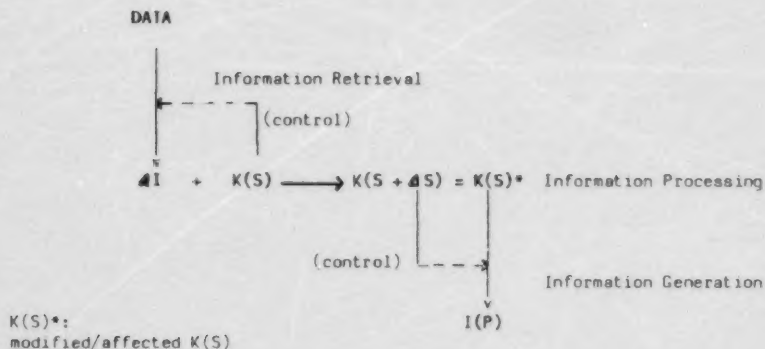


Fig 1b: Information Processing

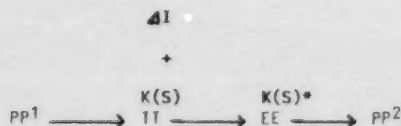


Fig 1c: Problem solving

In online IR the searcher operates on two kinds of data, communicated by the system:

- Conceptual Knowledge data
- System Setting data

On the other hand the IR system receives similar data from the searcher. The system's information processing, however, follows a slightly different pattern than outlined in fig. 1.

I is reduced to solely recognised information. Consequently, when rejecting non-recognised (= new, unknown information = data rest) present IR system's K(S) remain the same. However, it may look intelligent and perceptive, exactly because it displays objective knowledge structures containing ideas of searcher behaviour.

3. - KNOWLEDGE STRUCTURES IN IR

A searcher is defined as an intermediary, e.g. an information specialist or an end-user, i.e. the person or team that need to use information for example to solve a problem.

To be able to interrogate an IR system the searcher must possess sufficient "IR knowledge" consisting of knowledge on:

- the system setting
- the retrieval process

To be able to process the conceptual computer responses as information he/she must possess a certain degree of "conceptual knowledge".

When initiating a search the end-user has recognised an inadequacy in his/her conceptual state of knowledge, named ASK by Belkin and Ingwersen - see fig 2. - and The Formalised Need by Taylor (14).

If an intermediary mechanism is involved that person or machine must possess an ability to "picture" what the ASK is concerned with.

The systems total image of the world consists of the setting as well as the conceptual relations generated by authors. Thus, the notion "Generator" contains implicitly objective knowledge structures deriving from authors, systems designers and maintenance staff as well as file producers - all responsible for the total image of the IR system (15).

The model emphasises the possible implications and complications to be considered in designing "intelligent" IR systems, eventually replacing the human intermediary:

1) the dynamic interaction between conceptual knowledge, actual perception of needs and professional IR knowledge; 2) the limitations in the interface situation of the involved knowledge structures with respect to transformation of data into valid information (3).

Fig. 2 - A cognitive communication system for Information Retrieval
[from Belkin (2) and Ingwersen (3)]

If a human intermediary participates, the crucial problems for him/her are to be able to:

- a) understand the user's ASK - forming a picture of it
- b) analyse the current IR situation and the setting of the IR system(s)
- c) create representations of conceptual KS of large scale systems in relation to the subject - to form part of his own conceptual KS
- d) link interactively the picture and the representation, presenting potential information to the user.

The intermediary may not be in a position of sufficient conceptual KS to deal with the subject asked for. The optimal situation in such cases is when the end-user and the intermediary act together - each one operating his own speciality. Klaséna survey of the Swedish online population from 1982 emphasises explicitly this problem area (16).

In case of automated intermediaries - whether forming part of a mainframe or being a personal computer - the problems concerned are basically similar to those of humans. It is the automated interrogation, analysis and restructuring mechanisms to which one has to pay attention. Consequently, mainframe and micro software should be designed to enable the involved knowledge structures to adapt their model of the others through a series of interactions, i.e. instantaneous, complex learning processes are involved. This relational software flexibility requires considerable working and memory space if it is to be applied to large scale IR systems, containing databases with up to 4 to 6 million records - e.g. PASCAL and CHEMABS on ESA-IRS. Relevance feedback features like e.g. ZOOM point in this direction.

If an intermediary mechanism does not participate the end-user may lack IR knowledge to such an extent that IR becomes a kind of gamble. Besides, he/she may be unable to "picture" his/hers own information need.

So far only few attempts have been made to produce experimental, automated intermediaries. Meadow et al. has created the IIDA package which serves as an IR training device and acts as IR supervisor during real online searching on DIALOG (17). IIDA operates only on b) - using diagnostics of the interface.

Preece has presented an online associative query modification system operating on network representation of controlled vocabulary and thesaurus in very small databases (18). This system is based on the "best match" principle and works on parts of c), not aiming at adapting to the user's ASK. The experimental ASK system by Belkin et al. relies mainly on data concerned with c) and d). It has so far not been applied to large scale systems and no automated device exists to obtain information about the user's ASK. Instead, the ASK statements are regarded as identical to the need "picture", deriving from a human interview (19).

4. - The SEARCHER's KNOWLEDGE STRUCTURES

Several recent studies of online user communities distinguish between different kinds of intermediaries and end-users. Klasén's investigations demonstrate the importance of IR skills and experiences (16). Howard shows how the interface activity is dependent on knowledge of search language facilities and database structures, i.e. the setting in fig. 2 (20).

From a cognitive point of view four basic searcher groups seem obvious at the event of solving a need:

1. - the "elite"
persons possessing both IR and highly specialised conceptual knowledge
2. - the "intermediary"
persons with IR knowledge but scarcely specialised conceptual knowledge
3. - the "end-user"
persons possessing highly specialised knowledge but scarce or no IR knowledge
4. - the "citizen"
persons having neither specialised or IR knowledge in general

The elite group is small - approximately 15% (16) - and consists of mainly information specialists who may act as intermediary or end-user within their specialised subject field. As indicated by the survey of UK online users, a large amount of searches can be attributed to this group - perhaps up to 40% (21).

Group 2 may be regarded as the typical intermediary group of today. It consists of librarians, information specialists and others who have obtained IR skills via formal education and/or experience.

All online user studies estimate that a vast majority of today's searches are carried out by intermediaries. Klasén, for example states 75%. In his survey, the notion "intermediary" refers to group 2 and parts of group 1 and 3.

The typical end-user belongs to group 3. No surveys cover as yet group 4. These persons are the potential future users on the online market: the ordinary citizen. At present it seems that the end-user group is growing faster than the others.

The retrieval result depends on whether there is or may be created accordance - via transfer of data into information - between the searcher's KS, i.e. IR structures and conceptual KS, and the generator's KS, i.e. the setting and the system's conceptual KS - see fig. 2.

The "elite" searchers are the most fortunate ones. They possess a maximum of IR and conceptual knowledge. According to Howard very experienced searchers with knowledge of the field have the lowest preparation and connect time and highest use of commands, terms searched, sets viewed and the best results (20). This may indicate that this group is quite efficient to adapt to (learn) the setting and to process the conceptual KS of the system in relation to ASKs.

Klasén's investigation shows some interesting characteristics regarding group 2 and 3.

Group 2 accounts for approximately 50% of the online community, all actually being intermediaries.

Group 3 accounts for 35% - out of which 55% are intermediaries and 45% end-users.

Individual intermediaries (in Sweden) search 1,3 hours per month, with knowledge on approx. 4 databases and processing 2 files per search. Within 1/2 year they apply 2,5 different IR systems, i.e. 2,5 command languages and other system's settings.

Individual end-users search 0,5 hours monthly, being familiar with approx. 2,3 databases and applying 1,8 database per search. They use less than 2 host systems half yearly. (16)

As may be seen many searchers, especially from group 3, are so called "casual users". They show individual search frequencies far too low to keep up with changes in the search language facilities.

Search results become then more expensive and less valid than if carried out by means of other commands.

In view of the present state of increasing complexity of IR systems with respect to the contents of overlapping databases, the rapid growth of document representations and full text items as well as the growing complexity of search languages, casual searching is bound to disqualify itself.

It is up to the systems really to support the searcher in the event of interface, presenting more simple, flexible and automated (intelligent?) facilities. Otherwise, the situation will imply that only an elite group can be trained and operate the IR systems qualitatively, thus inhibiting online penetration to a mass market, as commented by the UK survey (21).

5. - The COMMAND LANGUAGE and INFORMATION

Each command language consists of commands and sub-commands that constitute a consecutive range of search facilities. The complexity of IR, the number of features and the nature of information needed make it extremely difficult algorithmically to establish a "natural" order or sequence. However, it is possible to chunk certain individual command sequences automatically.

Although the rules for command execution are more relaxed than in the past, online IR requires exact knowledge on the syntax of the single language. As a natural consequence of the system's fixed setting it is assumed that searchers always know of the impact of correctly entered search statements. Furthermore, the design presents at a maximum only explanations in relation to (parts of) the setting. In case of incorrect use of commands and/or rules the computer replies are far less informative than replies based on correct input. Today's error messages are inadequate and non-refined, designed to support searchers with a vast IR background - group 1 and 2, section 4.

With the broadened scope of searchers in mind the online industry has very recently initiated some new ways of applying information technology to the interface situation.

The following approaches - sometimes contradictory often complementary - are visible:

- the traditional
 - a) amending new command facilities
- the instructive
 - b) intensifying the application of command and search instruction
- the automated
 - c) automating the IR process on semi or full scale

Almost all hosts apply the traditional view, which is needed to meet the demands for specific searching. This was the reason for introducing free text operators and to a certain extent also crossfile searching. However, each albeit necessary improvement increases the effort needed to interpret the impacts of the new facilities. Casual users, beginners and potential customers are in weak positions. Although search motivated they seem to stick to basic routines and commands and avoid more sophisticated ones (22). As indicated by the surveys, a great amount of system setting data never become information (16, 20, 21) - see fig. 1a. Potential users see the languages as a barrier.

The instructive - or the IR knowledge - approach opens up a new spectrum. It is mainly caused by the microcomputer environment which in two aspects imposes a challenge to the online industry.

Firstly, several independent micro training packages dealing with online command languages and searching are put on the market. SEQUOL is a simulation program applied to ESA-QUEST consisting of modules and test quiz (23). Other packages deal with both commands and the search process (24). They rely on diagnostics and contain more or less informative instructions in case of errors. They derive from the decade long experiences with Computer Aided Instruction (CAI).

Secondly, intermediary or expert systems emerge (17). Applied during current searching they offer both training and actual IR instruction. However, the complexity of IR make their diagnostics simple and didactically insufficient for the time being.

One result of this approach is to put online training programs on mainframes and to convert diagnostics and instructions into the host computers. Thus, the command language is augmented with more informative computer replies - whereby the demands of the different user groups can be satisfied. By operating on several levels of IR knowledge it is assured that experienced searchers don't get bored and beginners perceive sufficient information.

The automated approach is only in gestation in relation to operational IR systems. Again, the micro evolution may rapidly change the picture. Download of software, setting data and conceptual data from mainframes to extended micro computer memory spaces is a fact today (25). What is lacking behind this technological enhancement is our full understanding of the IR interface, as stated in relation to fig. 2. It would seem a promising experiment to apply the ASK mapping model (19) to a database extract, downloaded from a mainframe in relation to real searching.

What is already becoming operational is semi-automation applied to the setting. Basically two kinds of command features exist: the "multi-option" type, containing several parallel, independent functions and the "sequential" type, which - to fulfil their aim - consists of consecutive functions. Only the latter, or combinations of the two, may result in automation and minimisation of commands.

Free text searching using positional operators is a typical example of the "multi-option" kind of facility. The FIND command - incorporating SELECT and COMBINE functions - and QUESTINDEX are "sequential" features. ZOOM shows characteristics of both types.

5.1 - Free text searching

=====

Communication and information requests are normally made in natural language and in context.

Free text searching offers in principle options by which the searcher may search more specifically, i.e. on broad contexts, expressions and composite words from titles, sources, descriptor fields and most important: from the abstracts.

It is the aspect of contextualisation which make free text searching attractive, since it decreases the degree of concept ambiguity. It ought to improve relevance because searching and retrieval of a context may assure that the retrieved text means the same to both author and searcher. Thus, "potential information" becomes easier useful "information" - see fig. 1 and 2.

Free text operations are carried out by means of contextual logic using positional operators, truncation and masking. In short they allow for searching on adjacent words, words with up to 5 words distance in fixed order, words or expressions placed in the same sentence, searchable field or citation. Each option has its individual operator. Unfortunately for searchers the operators are designated differently from system to system. Adjacency may for example be communicated correctly to ESA-QUEST, CCL and BRS using respectively: (W) - "space" character - ADJ.

to use the operators in full scale requires a vast IR knowledge background. The impact of the different options must be well known before-hand and linguistical elements like spelling, singular/plural forms, common/scientific language and language syntax, synonyms.... well understood.

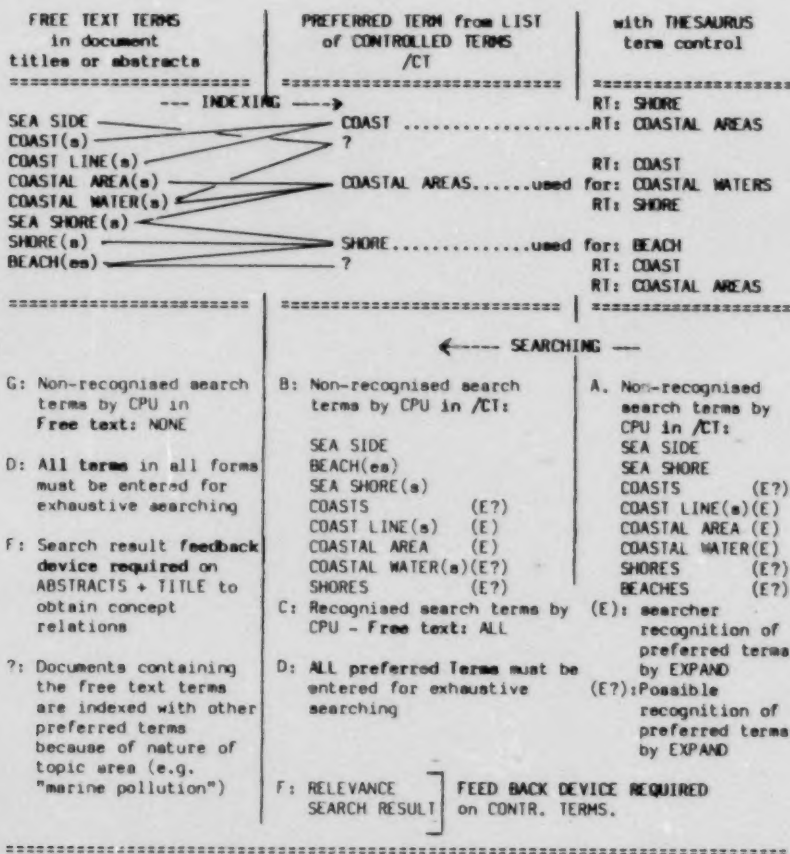


FIG 3 Simplified example of searching on conceptual KS in database via thesaurus (right), via controlled terms alone or free text (middle) and only via free text (left)

Although extremely adequate in IR - see fig. 3 - nearly all online search studies show a rather low application frequency of free text. This may also be due to the fact that a great number of surveys are based on the ERIC database and alike well equipped files which are rare on the market (20, 26).

As illustrated in fig. 3 and stated by studies of search outcome in relation to controlled vocabulary and free text, the only way to maximise the return is to apply both controlled and free text terms (27).

The problem then is the lack of online support in the interface situation when applying free text searching. A semantic gap exists between author and searcher language. To come up with synonyms and/or closely related concepts and other linguistic elements within a subject area requires conceptual analysis in depth (28). If no controlled vocabulary exists in a file and no feedback mechanism is available on the system, the outcome depends solely on the searcher's conceptual and setting knowledge - fig. 3 left.

If controlled terms are present in the file, free text searching will include retrieval of such terms. Because of the aids created by designers and indexers, the searcher needs IR knowledge on how to enter and use them - fig. 3 middle, right.

To search on controlled vocabulary (and thesaurii) alone is conceptually easy but requires a certain amount of IR knowledge. It applies to intermediaries and some end-users possessing IR skills. However, the outcome is only improved by use of free text operations. This requires developed IR and conceptual knowledge, normally only possessed by a smaller elite - group 1. Functions most frequently used are those related to adjacency and truncation - other operations are rarely applied or avoided. "Quick and dirty" searching is applied more often than one should expect (22).

5.2 - Crossfile searching (QUESTINDEX)

=====

As for free text searching this facility is designed within the framework of the traditional view. In the ESA-IRS version some automation is built into the feature. Its aim is to support the searcher in the choice of valid databases and additionally to indicate the validity of conceptual search statements in respective files. Crossfile searching is a typical spin-off effect of a rapidly growing information system.

However, like all other operational interface facilities this option is not prompted to the searcher, for example at initiation of a search, even at least two files normally should be searched for exhaustivity reasons. More "instructive" command languages, e.g. on microcomputer intermediary systems or future mainframes, will hopefully include the prompt and automatically perform the crossfiling.

Opposite free text searching QUESTINDEX is sequential in its scope. Basically it consists of the following steps:

1. to enter the facility (into e.g. a separate file)
2. to enter a list of standard topic areas that contain predefined files
3. to choose a topic area (=group of files)
4. to enter the search statement to be crossfiled automatically

On ESA-QUEST the search statement may consist of a single term, a free text context or an entire saved search.

The simplest crossfile search is the QUICK QUESTINDEX whereby the searcher enters his search statement preceded by a "Q" or "QUEST", e.g.

QUEST FIND INFORMATION RETRIEVAL AND MICROCOMP?

Since no separate file exist on ESA-IRS it is the "Q" which assures that the cross-searching is performed. (Forgetting to enter the "Q" results in execution in the current file). The crossfile searching is now performed automatically in all multidisciplinary IRS databases, i.e. approximately 16 million records are scanned in seconds by the computer. This basic option reduces the steps to one - step 4. More important to casual or potential users, the automation adds only one command (incl. impacts) to the searcher's IR knowledge structures.

At all, QUESTINDEX requires the knowledge of 3 commands if performed correctly:

- | | |
|-------------------------|--|
| QUEST TOPICS? | - displaying list of standard topics |
| QUEST TOPIC INFORMATION | - choosing a topic, e.g. "information" |
| Q FIND "statement" | - to enter search statement |

From a cognitive view two aspects are of importance:

1. The IR knowledge required to enter and to use the facility
2. The conceptual knowledge required

To enter the feature. Performing quick cross-searching the common search commands must be current knowledge. To select a valid topic and to perform the search across the databases requires in addition an ability to separate the two decisive, but similar commands from each other. Furthermore, one must never forget the preceding "Q". Without computer prompts only full time searchers frequent the facility sufficiently to know it by heart.

To select an appropriate topic area from the standard list requires that the ASK need statement(s) is sufficiently categorised to be related to one of the topics. This standard index design problem in librarianship is among others dealt with by Keen in his investigations on information processing of printed subject entries during searching (29). The problem of matching is in line with situational and categorial classifications of objects as described by Ingwersen (3). At present the number of topic headings in alphabetical order are small - 30 to 40, i.e. two screen pages at maximum. It does not seem a difficult task for all searcher groups to relate their abstract or situational ASK statements to a heading.

To use the feature. Output data from the computer contains database name and number, timespan and number of postings related to the entered search statement for each file.

This implies that the searcher only gets new information concerned with the figure and perhaps with the file name/number. No indications of file size is given and it is thus evident that he/she has to possess knowledge on subject coverages, indexing policies, file structures and sizes, to be able to use the computer reply. Otherwise the output remains data or contains very limited information value.

In this case the file number may be of use for file switching - but first when the posting figure is interpreted. The example below will demonstrate the problem and in addition show that common linguistic aspects, discussed under free text searching, are involved:

If a searcher crossfiles for example the statement "micro computer" in the standard "computer" files, even in free text mode, and obtains few or zero postings in file A and B, he may avoid using them.

If A and B are the INSPEC and NTIS databases, can he afford to eliminate them? In fact the result is very few or zero postings dependent on the mode of searching. Among 7 to 10 cross-searched databases, they might be avoided, if their names do not trigger any part of the searcher's database knowledge. They become information leading to possible wrong decisions made by the K(S)* - see fig. 1a-c.

Today, crossfile searching requires a vast IR background on system setting and the retrieval process as well as conceptual knowledge, for example on linguistic matters as spelling, terminology.... Only if possessing the proper KS (the elite searcher), one is able to use the feature to select valid databases and to judge the relevance of the statement(s) presented to the computer. For group 2 intermediaries it becomes merely a tool for file selection and only to some extent an aid as term relevance feedback mechanism.

The end-user with scarce IR knowledge may apply it to choose files. However, he may choose the wrong ones, if a valid file does not (or rarely) contain the search statement. The identical gamble is the concern of the layman.

5.3 - Term frequency analysis (ZOOM)

=====

ZOOM is a consequence of both free text and crossfile searching. Although the feature is a typical "multi-option" facility, some automation is incorporated.

ZOOM is designed to analyse the frequency of single words, phrases or codes appearing in a selected set of references. It is based on the converted versions of the large scale IRS databases. For each database a selected number of searchable fields are made analysable. The following fields - common to all files - can be zoomed: Titles, Authors, Corporate sources, Abstracts, Classification codes, Controlled and Uncontrolled terms (=Index terms). In certain databases additional special fields are zoomable, e.g. Molecular formulae and Registry numbers on CHEMABS.

ZOOM was invented and applied to the ESA-QUEST search language in order to improve and support the casual user's search performance (30). By term frequency analysis of a selected SET of references the outcome is a ranked list of designations of concepts which in one way or another are associated with each other within the selected pool of documents.

Hence, ZOOM may support the searcher in several ways: It may

- * suggest potentially useful, new search entries
 - display preferred index terms (see fig. 3, middle)
 - suggest new terminology and synonyms (see fig. 3, left)
 - point out potential related conceptual aspects
 - allow for specific searching
 - show potential natural language concepts (see fig. 3, left)
 - display different spelling (UK/US) and similar concepts in other languages
- * give ideas and associations in relation to precision of the ASK - near mis solutions
 - make the searcher discover ambiguities and homographs
- * work as a relevance feedback tool for conceptual contents of the final set of documents

Besides being applied to an operational IR system ZOOM is unique in the sense that it may operate within files consisting of pure free text terms - see the requirement on fig. 3 left.

The ZOOM facility may also act as an online bibliometric instrument, available to researchers and file producers who already use it to improve controlled vocabularies by discovery and introduction of new terms.

The frequency analysis is by default carried out on the basis of 50 records across the last selected set. The computer working time is thus minimal and tests shows that 50 records are enough to obtain relevant results.

By default the analysis is made on the index terms.
In its simplest form the command

ZOOM or Z

thus implies frequency analysis of all index terms - phrases or single keywords - applied to 50 records across the last set. This command results in the feedback device required to obtain potential preferred terms associated with a search statement, as mentioned in fig. 3 middle and right.

Other command possibilities are for example:

- ZOOM TITLE WORDS - analysis of single title words from 50 records across the last set
- ZOOM LATEST AUTHORS - analysis of author names of the latest 50 records in last set
- ZOOM 4 (200) LATEST - analysis of the index terms of the 200 latest records of set 4

Below are shown some examples of ZOOMing. This is followed by a discussion of the impacts in relation to the searcher's KS. Finally some enhancements regarding the interface situation is made.

5.3.1 - ZOOM samples

=====

To fill the semantic gap between the searcher's KS and that of the system, one could - prior to ZOOM - look up in the expand list of controlled or free text terms - or display few records, scanning the title, abstract and index term fields.

Let us reconsider the QUESTINDEX result on "micro computer" in INSPEC (It is assumed that INSPEC is known to contain in depth coverage of electronic engineering and consists of 1.9 million refs.)

First we enter the search argument in free text mode:

FIND MICRO COMPUTER? ? - CPU reply:

1 342 MICRO(W)COMPUTER? ?

Then we ZOOM set 1:

Computer reply:

	Text Analysis Results
Freq	Words/Phrases
16	MICROCOMPUTERS
7	MICRO COMPUTER
6	MICRO COMPUTERS
6	MICROCOMPUTER
4	COMPUTER NETWORKS
4	COMPUTERISED
	INSTRUMENTATION
4	MINICOMPUTERS
4	SYSTEMS ANALYSIS
3	AUTOMOBILES
3	COMPUTERISED CONTROL
3	DATA PROCESSING
3	MEDICAL DIAGNOSTIC
	COMPUTING
3	MICROCOMPUTER
	CONTROL
3	PERSONAL COMPUTING

...Pages.Lines: More=20.16

Because of the high frequency of the preferred terms - e.g. "microcomputer(s)" - it is rather easy even for beginners or persons with scarce subject knowledge to pin-point the adequate terms and add them to the search.

In addition, some associated terms may be grasped - if their conceptual relation to the search argument is known to the searcher. For example the term "personal computing" which may trigger the concept "personal computer(s)" by memory recall.

Synonyms can be spotted and more precise ASK expressions obtained whereby ambiguity can be avoided, via ZOOM:

We consider a situation in which the ASK is about "body factors under zero gravity conditions in space experiments". It is, however, expressed in more general terms by the user, e.g. "man and zero gravity in space". We assume that the end-user/intermediary tries the NASA file.

Prior to ZOOM he might have entered the "quick and dirty" strategy: SELECT MAN; SELECT GRAVITY; COMBINE 1 AND 2 to be sure to obtain references. With ZOOM in mind, specific searching in free text can easily be performed with a good result:

FIND MAN AND ZERO GRAVITY - reply:

```
1 9240 MAN
2 399 ZERO(W)GRAVITY
3 6 C 1*2
```

Considering the size of NASA (over 1 million records), its space science coverage and the impact of the entered concept, one may decide to supplement the argument.

ZOOMing set 3 gives the following list in shortened form:

Frq.	Word/phrases
6	WIGHTLESSNESS
2	GRAVITY
2	HUMAN BODY
2	HUMAN FACTORS ENGINEERING
2	MAN
2	MOTION
:	:
:	:
2	ZERO
1	ANIMAL

As in the former example another preferred term seems relevant: "weightlessness". As a synonym to "man" "human" seem obvious. Perhaps the term "human body" catches the eye of the user and maybe he wishes to exclude "animal(s)" from the search. His ASK may very well be more elaborated during this learning and processing process - see fig 1 b and c - e.g. "human body factors under zero gravity or weightlessness conditions in space, human experiments"

After having manipulated and combined sets he may finally apply ZOOM as a relevance tool - not yet displaying any records:

```
Frq Words/Phrases
-----
46 HUMAN BODY
41 WEIGHTLESSNESS
27 AEROSPACE MEDICINE
18 SPACE FLIGHT STRESS
16 PHYSIOLOGICAL
    EFFECTS
10 PHYSIOLOGICAL
    RESPONSES
10 WEIGHTLESSNESS
    SIMULATION
 7 BLOOD CIRCULATION
 7 CARDIOVASCULAR
    SYSTEM
 7 HEMODYNAMIC
    RESPONSES
 7 HYPOKINESIA
 6 BED REST
 5 ADAPTATION
...Pages.Lines: More: 4.2
```

(The sample is from a search on "human body" * (zero gravity + weightlessness) - animal(s))

.....

5.3.2 - Discussion and enhancements of ZOOM

=====

A high frequency will in general indicate a close relationship to the search argument in the zoomed set or demonstrate main topic areas within the field of interest. Preferred terms which may make supplement to the searched one(s) are normally placed on the top of the list.

This make ZOOM quite relevant to apply for all user groups, especially intermediaries without special knowledge of the subject.

Low frequencies must not be overlooked. It is down on the lists that sub-areas and aspects are hidden - not only in relation to the executed arguments but also associated with the other concepts displayed. If one zooms the title and abstract fields one may discover new terminology and get conceptual surprises.

Actually, zooming on one or two terms in fig. 3, left gave nearly all the rest. After restructuring a search a new frequency analysis across abstract and title words resulted in all relevant terms plus some new associations.

However, as special subject idea bank, ZOOM only serves conceptually specialised persons.

Besides the basic and simple ZOOM command a great number of command combinations exist. To grasp the impact of them all is a severe task.

As for other traditional interface facilities the computer reply does not contain sufficient setting data. At present ZOOM is in a test period where users can suggest improvements.

Most important is the layout design of the replies. Data concerned with the database in which the analysis is carried out ought to be shown automatically together with the analysis results. Likewise for the analysed set number and the analysis parameters. At present it is necessary to prompt the computer to give this basic information.

It is the recognition facets which make ZOOM such a powerful interface mechanism on a operational system. As stated earlier in relation to fig. 1, recognised concepts may carry with them new, related ones or trigger memory recall activity.

It is the receiving searcher's KS which must establish the valid associations. An improvement might be to incorporate online thesaurii when available on the specific files. The CALIBAN system described by Frei and Jauslin operating on a personal computer with tree structured displays of thesaurus terms, may indicate such a possibility (31). However, most files do not contain thesaurii at all and for full text systems in natural language elaborated zooming combined with weighting of search terms might seem promising.

Another enhancement might be to combine mapping of user ASK statements and weighting matched with ZOOM representations of document texts. However, all this is future on commercial systems.

It is more likely that automation of existing command sequences will emerge. For example auto file select combined with automatic merging of records and algorithmic zooming and displayed alternatives.

The download perspectives may introduce new microcomputer solutions including frequency analysis on downloaded material (25). Parts of the analysis may be automatically retransmitted, signifying a kind of corrigation "map" on which a modified search can be performed.

7. - CONCLUDING REMARKS

Present IR systems are designed almost only for IR experienced end-users and intermediaries with more than common conceptual knowledge.

The command language is the key - but often the searcher is only equipped with the "kitchen" key which does not fit the yale lock.

A high amount of data is not made available or never becomes information. Many computer replies only contain data of limited information value.

The scope of command languages and other interface devices ought to be broadened up. This does not imply that the system continuously must prompt searchers or signal the next step, explaining all in detail. That would be too boring to a great part of the user community. To broaden the scope means to flexibilise the facilities; to make available by default more space for instructive messages and simple commands to beginners and casual users.

More advanced searchers may switch them off - if wished so. This signifies a step forward towards "learned information systems". The technology is ready. If not on mainframes then on micros. It waits to be used to make the tools your own.

ACKNOWLEDGEMENTS

The author wishes to thank the Information Retrieval Service of the European Space Agency for its collaboration and support. Special thanks are given to the colleagues of IRS who have all contributed in developing the framework of this paper, in particular, Dr. Georges A. Proca, Head of Online Services Division.

REFERENCES

1. DE MEY, M. The relevance of the cognitive paradigm for information science. Theory and application of information research: Proceedings of the SIRE 2 conference, 1977, Copenhagen. / ed. by Ole Harbo and Leif Kajberg. London: Mansell, 1980, pp. 48-61
2. BELKIN, N.J., ODDY, R.N., BROOKS, H.M. ASK for information retrieval: Part I. Background and theory. Journal of Documentation, Vol. 38, no. 2, 1982, pp. 61-71
3. INGWERSEN, P. Search procedures in the library analysed from the cognitive point of view. Journal of Documentation, Vol. 38, no. 3, Sept. 1982, pp. 165-191
4. DE MEY, M. The cognitive viewpoint: its development and its scope. International Workshop on the Cognitive Viewpoint, 1977 / ed. by M. De Mey, University of Ghent, 1977. pp. xvi-xxxii
5. BROOKES, B.C. Discussion of knowledge spaces related to Popper's 3 World Model. Proceedings of the 3rd International Research Forum in Information Science (IRFIS 3) / ed. by Tor Henriksen. Oslo, 1979. p. 254 (Workshop 4: Information and Information Carriers)
6. MACKEY, D.M. What makes a question. The Listener, 63 (May 5, 1960), pp. 789-790
7. BROOKES, B.C. The foundations of information science, part I: Philosophical aspects. Journal of Information Science, 2 (3,4), 1980, pp. 125-133
8. SHANK, R.C. The role of memory in language processing. The structure of human memory / ed. by Charles N. Cofer. San Francisco, 1975. pp. 162-189
9. NORMAN, D.A. and BOBROW, D.G. On the role of active memory processes in perception and cognition. The structure of human memory / ed. by Charles N. Cofer. San Francisco, 1975. pp. 114-132
10. NOVAK, J.D. Concept-based learning. Presented at the Amherst Symposium on the Optimum Utilisation of knowledge, Amherst, Mass. Nov. 5-8, 1981, 22 p.
11. GOWIN, D.B. The structure of knowledge. Educational Theory, vol. 20, no. 4, 1970
12. MERLEAU-PONTY, M. Le visible et l'invisible. Paris: Gallimard, 1964
13. POPPER, K.R. Objective knowledge: An evolutionary approach. Oxford: Clarendon Press, 1973
14. TAYLOR, R.S. Question-negotiation and information seeking in libraries. College and Research libraries, 29(3), 1968. pp. 178-194
15. HARBO, O, INGWERSEN, P., TIMMERMANN, P. Cognitive processes in information storage and retrieval. International Workshop on the Cognitive Viewpoint, 1977. / ed. by Marc De Mey, University of Ghent, 1977. pp. 214-218
16. KLASÉN, L. Online-användning. Rapport. DFI, 1982. Stockholm: Delegationen för vetenskaplig och teknisk informationsförsörjning, 1982:1 (80-128). ISSN: 0349-9324 (Engl. title: Online use)

17. MEADOW, C.T. et al. A computer intermediary for interactive database searching, part I: Design. *Journal of the American Society for Information Science*, Sept. 1982, pp. 325-332. Part II: Evaluation. *JASIS*, Nov. 1982, pp. 357-364.
18. PREECE, S.E. An online associative query modification methodology. *Online Review*, Vol. 4, 1980, pp. 375-382.
19. BELKIN, N., ODDY, R.N., BROOKS, H.M. ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation*, Vol. 38, no. 3, Sep. 1982, pp. 145-164.
20. HOWARD, H. Measures that discriminate among online searchers with different training and experience. *Online Review*, Vol. 6, no. 4, 1982, pp. 315-327.
21. ONLINE INFORMATION CENTRE. Survey of online users in the United Kingdom. ASLIB, 1983.
22. WORMELL, I. Computerised statistical research study on ESA-QUEST command use: Interim Report. Unpublished. ESA-IRS, 1981.
23. SCOLTOCK, J. SEQUOL: A simulation of ESA-QUEST on-line. University of Sheffield, 1982. 199 p. (Thesis).
24. WALDSTEIN, R. DIATOM: A DIALOG simulator. *Online*, 5(3), 1981, pp. 68-72.
25. ONLINE SERVICE DIVISION. ESA-IRS. Download. *News and Views*, Vol 8, no. 4, 1983, pp. 25-26.
26. MARKEY, K., ATHERTON, P., NEWTON, C. An analysis of controlled vocabulary and free text search statements in online searches. *Online Review*, Vol. 4, no. 3, 1980, pp. 225-236.
27. CALKINS, M.L. Free text or controlled vocabulary? A case history step-by-step analysis ... plus other aspects of search strategy. *Database*, Vol. 3, no. 2, 1980, pp. 53-67.
28. DOSZKOC, T. E. Implementing an associative search interface in a large online bibliographic database environment. *New Trends in Documentation*. 39th FID Congress, Edinburgh, 1978. ASLIB, 1980, pp 295-297.
29. KEEN, E.M. On the processing of printed subject index entries during searching. *Journal of Documentation*, Vol. 33, no. 4, 1977, pp. 266-277.
30. MARTIN, W.A. Helping the less experienced user. *Proceedings of the 6th International Online Information Meeting*, 1982. Oxford: Learned Information, 1982. pp. 67-76.
31. FREI, H.P., JAUSLIN, J.-F. Graphical presentation of information and services: A user-oriented interface. *Information Technology: Research and Development*, Vol. 2, 1983 pp. 23-42.

USING MONSTRAT FOR THE ANALYSIS OF INFORMATION INTERACTION

N.J. Belkin,

City University London;

G. Windel,

Projekt INSTRAT, Free University Berlin

1. INTRODUCTION

Wersig and Hennings, in this volume, have presented an overview of the general MONSTRAT model of the functions performed by the ideal information provision mechanism (IPM).

That description was largely theory based and descriptive.

In this paper, we intend to demonstrate that the MONSTRAT model can be used analytically, and in doing so, to show how it can be applied empirically. We do this by reporting on several empirical investigations of human/human information interaction of different types, which used the MONSTRAT functions as the analytic tool. The goals of the investigations themselves will be specified in our descriptions of them below, which are necessarily sketchy. We wish here to use our results: to demonstrate that the MONSTRAT functions do indeed appear in information interaction and appear to be necessary to "successful" information interaction; to indicate how the MONSTRAT approach can be used in IPM improvement and design; and, to indicate how separate investigations of information interaction, from various points of view, can be integrated within a single framework.

2. DIALOGUE ANALYSIS

We will describe three different investigations of human/human information interaction which have been carried out by us within the MONSTRAT framework. They have each entailed some kind of dialogue or discourse analysis, but differ in the level and type of analysis and the in materials being analysed. We will refer to these investigations as: "interpersonal interaction"; "problem description"; and, "on-line functions".

In this section we describe the methods used and the data analysed in each investigation, and in the subsequent section some results of each of the investigations.

2.1 Interpersonal interaction

This investigation is reported in detail in Progris (1). Its goal was to investigate the nature of user/intermediary interaction in on-line interviews, especially in terms of the social and role constraints that influence that interaction, in order to get a general, descriptive account of what is going on in such interaction. This was largely exploratory research aimed at seeing if the MONSTRAT functions appear in information interaction, and if not, why not. The method was both participant and non-participant observation of eight on-line interviews and searches, which consisted of an interview between a user (sometimes the investigator) and a human intermediary, at which users were sometimes present. The data which were analysed were informal protocols, recorded by the investigators, of the interaction.

2.2 Problem description

This investigation was part of a project to consider in detail the problem description function of the MONSTRAT model. It was reported in detail in Belkin, et.al. (to appear). (2)

Its goal was to analyse a number of information interactions in order to isolate performances of the problem description function by intermediary, and then to subcategorize this function and determine the resources and procedures used by the intermediary to accomplish it. The eventual goal of the project was to specify the problem description function sufficiently for simulation in a computer system. The data analysed were annotated records of various sorts of information interactions, primarily advisory, which took place in real settings. These interactions were taped recorded, and the parties to the interaction were interviewed to determine their opinions about the interaction. These combined data were then transcribed and analysed utterance by utterance according to the MONSTRAT functions, with each instance of

problem description being further analysed. At the present time, the full analysis is available for only two interactions.

2.3 On-line functions

This work is the result of a project at the City University, London, reported in Brooks & Belkin (3), which was concerned with identifying and characterising the use of subject knowledge by the intermediary in the pre-search interview for on-line bibliographic retrieval. The motivating factor behind this work is the attempt to develop an 'expert intervention system' which might replace the human intermediary in some on-line searching situations, allowing direct end-user interaction. The data collected were tape-recordings of the interaction between user and intermediary at an academic on-line IR-service at the University of London, from the user's entry to the system to the end of the search. Two such interactions have been fully transcribed and analysed. The analysis is according to performance by each party of the MONSTRAT functions, each intermediary utterance also coded, as required, with the knowledge type was necessary for the intermediary to have made the utterance.

3. RESULTS

3.1 Interpersonal interaction

We now try to integrate the empirical findings concerning the role of the intermediary in online-retrieval processes into the complex MONSTRAT-model. First, we will deal with the MONSTRAT-functions which are performed by the intermediary in a specific way. The following statements rest upon the results of the interpersonal interaction.

In this study, several online retrieval intermediaries in different institutions were observed while working, and interviewed. Additionally a number of query processes were simulated by members of the research team and external test persons. (1)

3.1.1 The functions which appear

Input analyst

Input analyst is the function which actually understands the input - it isn't supposed to analyse it in any deep way, but rather just to know what has been said.

Usually, the input is made by written or verbalized queries. The intermediary doesn't analyse the input in a systematic way, but rather prefers to make an analysis by derivation from the user model. Input analysis, therefore, may be in many cases seen as some sort of a by-product taken from the reflection of another function.

Problem mode

In fulfilling this function, the intermediary is largely dependent on system quality, i.e. in our case reference retrieval. The nature of the problem or the respective state of problem treatment can only be taken into account if permitted by the abilities of the system. But what the intermediary can do is to "switch on", i.e. to refer to other reference systems or to human experts.

Problem state

Normally lacking ongoing interaction with the user the intermediary is not able to fulfill this function in the way discussed by the MONSTRAT-model. Instead, s/he usually draws his conclusions concerning the user's need situation by referring to the query. If that strategy fails, the intermediary takes the role of the "generalized other", i.e. the user to make a status analysis. In this case, his or her experience may help him.

As may be seen from these statements, there isn't such a thing as an interactive problem treatment process. Because this is so, important questions cannot be answered or even raised, like

- how important the problem is for the user
- which degree of urgency its solution has and
- where the consequences of the solution lie.

User mode

The predominant analytic elements which the intermediary draws on to depict an image of the user are socio-demographic criteria and his own experiences. Lacking continual interaction with the user and relying on experiences once made, the intermediary stereotypes the user. Leaning heavily on "every day psychology", the intermediary puts the user into a pigeon hole, which, taken the mentioned situation, may be often unavoidable.

Problem description

Performing this function, the intermediary tries to depict an image of the problem by building a relevant world with the help of generating descriptors and documents.

This is done intellectually, i.e. in the intermediary's mind. Changes and shifts may occur at this stage which do not relate to the problem itself, but to the intermediary's views and specific work strategies.

Similar to most of the other functions, the intermediary here too, is dependent on his own experience, knowledge store, and the representational potentialities of the system, though it is possible that he may get support from human experts.

In this context we have to bear in mind that the process of building an image of the problem is closely related to its verbal transcription, i.e. the intermediary, by depicting a problem image simultaneously is generating the respective descriptors for the query process.

Relevant world builder

The relevant world is built by selection of certain data pools, descriptors and their connection. Sometimes human experts are drawn upon, too. In most cases some sort of a model of this world segment already exists in the intermediaries' "own head", i.e. he knows or thinks he knows what the relevant world looks like.

In reality the relevant worlds of the intermediary and the system do not match, i.e. both "parties" have different segments of "world". Being aware of

the system's complexities and pitfalls the intermediary in some cases tends to neglect its potentialities by drawing upon his own relevant world straightforwardly.

Response generator

Performing this function, the intermediary is free to act as he wishes, i.e. he manipulates the system as long as the output seems to be appropriate for the problem and its treatment, which means that he selects and evaluates numerous documents.

Dialogue mode

This function is not performed by the intermediary. Instead, he restricts himself to adopting the results given to him by the system. This means, that various possibilities of communication and interaction are reduced to one mode, i.e. sending out the computer printouts. If the given answer is "understood" by the user, is up to the intermediaries' guess. Lacking a direct and issue based dialogue with the user, the intermediary has to take the risk of failure.

Output generator

Speaking in terms of the MONSTRAT model, the output generator converts the structures which the other functions have achieved into some form which the user can understand, in a way that is appropriate to the immediate situation. (4)

The reality of online retrieval processes gives us another picture. What the intermediary actually does is to give the user an "information package". Lacking a direct contact, the intermediary is normally unable to decide whether the given answer (i.e. references and abstracts) are understood by the user, in a way which is helpful to him.

World representation, is done primarily by episodic stores, which try to incorporate "world" via references.

An additional world representation is incorporated in the knowledge store and experience of the intermediary and other human experts.

World generation is performed by authors of documents, indexers and abstract authors. World generation, though, is limited by the system-induced-limitations (for example number of headings or lines for an abstract).

By printing out the log of the search and respective documents the dialogues between system and intermediary, as well as the dialogue within the latter is documented by what may be called a mechanism store. Through this procedure, the whole of all the other functions, though being not directly visible, nevertheless can be analysed.

3.1.2 The relationship of the functions to the social context

If we look at online-retrieval processes and ask for their interaction and communication related aspects from a social point of view, the results are rather meager.

As far as the social context is concerned, it is primarily the user analyst which is responsible for giving the intermediary a "picture" in which social and work related context the user and his problem are situated

- what kind of information he may need
- how much he may need and
- how urgently he may need the information.

In his role as input- and user "analyst", the intermediary is largely dependent on rather formal issues concerning the user's context. He usually doesn't know his concrete problem situation in all its relevant aspects. So by resuming his own experience with users, he tends to stereotype them. In that, he, for example, classifies the type of institution the user works in or his rank in the status hierarchy of that institution.

Interaction and communication, thus, is restricted from the beginning of the query process to a number of formal inputs, made by the user. From that moment on, the social context, in which problem treatment is dealt with, the intermediary is restricted to a man/machine relationship (with the exception of contacting other human experts).

This situation, where the social context is almost totally excluded and where the intermediary is left alone with the system and its potentialities, has grave consequences, such as

- only one mode (i.e. references) for all needs
- no taking account of problem state
- no user representation, i.e. user representation is not a part of the search process
- restricted possibility of problem representation.

Bound to the limitations of the system, the intermediary is forced to depict an image of the respective social context "within his own head". The relevant world builder can only give him a rather vague, filtered and indirect world representation, where the concrete world hardly shows up (mind, that the documents stored in the data bases are mostly written in scientific-technical terms and therefore only useful for a limited portion of their potential users.)

One of the eminent functions of the MONSTRAT-model is its high degree of interactivity, i.e., it must be seen as an interactive process between the information searcher and a mechanism, which provides the information. But in the hard realities of online-retrieval processes, interactivity as we have seen, is severely hampered by the inherent deficiencies of the user/-intermediary/system relation.

Dynamic functions like

- problem mode
- problem description
- dialogue mode or
- output generator

don't interact in a problem oriented form. Once and for all defined in terms of systematic requirements, they hardly leave any room for significant variation.

The, so to say, one-dimensional reduction of most functions can be seen in the

- formalized input
- existence of only one mode (references) for all needs

- not existing possibilities of problem analysis on behalf of the intermediary
- restricted possibilities of the problem representation
- rigid binding together of problem representation and relevant world
- severe restrictions concerning different choices by the answer model.

The intermediary's role in this process is an extremely difficult and may be even deplorable one. He has to fulfill a majority of the functions without being helped by the system or interactive processes with his client. He has to

- bring the input in a formatted form
- reduce the different needs to single modes
- make the different state modes equivalent
- develop a user model
- analyse the problem
- bring the problem in a system adapted representation
- construct a unity between problem representation and relevant world.

That means, that the deficiencies of the system are pressed upon the intermediary.

3.2 Problem description

Analysis by the problem description group led to two different types of results. In the first case, we concentrated upon identifying occurrences of the problem description function and then analysing these to try to infer exactly what processes and resources the intermediary (or advisor) used to accomplish the function (or sub-function).

This analysis was performed for two advisory interviews, and also for two interactions in a radio telephone advice program.

The results of this analysis of advisory interaction demonstrated that:

1. the problem description function was performed by the advisor consistently as a major part of the interaction.
2. the problem description function was multi-part, divided into subfunctions of: identification of problem type; specification of problem topic; specification of problem environment (or constraints); with a possible fourth

function, identification of problem context.

3. the advisor typically used problem description in a questioning mode, forcing interaction with the user.

From these results, we were able to infer the description of the problem description function indicated in figure 1.

PROBLEM DESCRIPTION	
TASK: Analysis of the problem (task) with respect to theme (issue), typ, structure	
ACTIVITIES	OBJECT
Analysis	Problem theme, - typ, - environment, - structure (cognitive)
Selection	Problem treatment strategies
Evaluation	
Generation	Hypothesis about the issue, the typ, structure and environment, problem description
Decision	
knowledge resource: Typologie of problem issues, cognitive problem structure and problem treatment strategies.	

fig. 1: Functions of the Problem Description

The second case in this study arose from the first, in an attempt to define the structure of the interaction between user and advisor in problem description. In analysis of one of the transcripts, we noted the intermediary followed a rather strict pattern of interaction with the user. This pattern turned out, on analysis, to be based on a structure, which the advisor held in her mind, of the nature of the problem.

Figure 2 shows the problem structure which we derived.

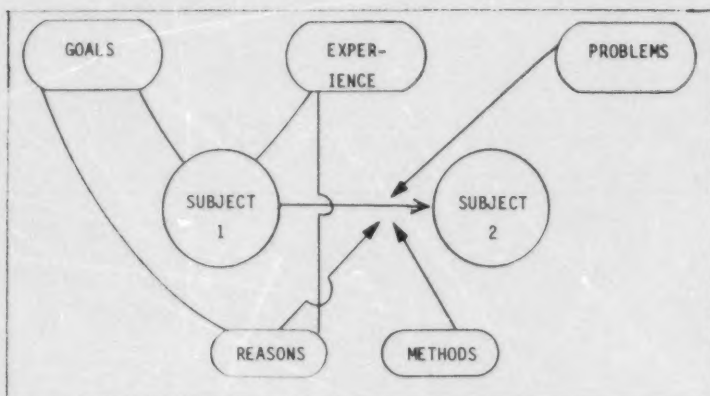


fig. 2: General structure for student problem type: change of major results of U.1 (in mind of I)

The problem description function was then seen to be related to this structure in specific ways, with different of the subfunctions being associated with different parts of the stereotypical problem structure. The advantage of this particular way of viewing the interaction is that it gives one a way to interpret the succession of functions (including others than the problem description function) in a regular way. The choice of stereotype structure appears to be based on the very early judgement by the advisor of what the pro-

blem type is, and our conjecture is that this is probably the case in all information interactions. Such a strategy is, of course, helpful, in that it limits the problem space for the advisor. On the other hand, unless commitment to the strategy is loose, using it may lead the advisor into very inappropriate interaction strategies.

3.3 On-line functions

In this work, we examined and coded in detail the transcripts of the pre-search interviews between user and human intermediary prior to an on-line search. In this case, the data were similar to those of the 'interpersonal interaction' investigation, but they were analysed in more specific detail.

The results of this work show that most of the basic MONSTRAT functions do indeed appear in information interaction, performed either by user or by intermediary, and they are also indicative of what makes a "successful" information interaction, in terms of occurrence and pattern of functions. The details of the analytic rules are presented in figure 3, and the records of function performance for user and intermediary over time in two interactions in figure 4-7. Perhaps the most important characteristic of these data that one should note is the discrepancy in functions performed between the intermediary of interaction 4 and the intermediary of interaction 5. Note also that in interview 4 there is a high degree of overlap of functions between user and intermediary, particularly in the problem description function. Although evaluation of the success of the respective interactions is not unambiguous, there is some evidence to suggest that interaction 4 was more successful than interaction 5, and we tend to the view that this can be explained by the different functional patterns on the part of the intermediary in the two interactions.

PROBLEM MODE (PM)

1. ASK
2. OFFER
 - A. MECHANISM'S CAPABILITIES IN GENERAL
 - B. ALTERNATIVE MECHANISMS

USER MODEL (UM)

1. ASK
2. OFFER
 - A. PERSONAL BACKGROUND (STATUS, ETC.)
 - B. INTENTIONS - GOALS
 - C. FAMILIARITY WITH INFORMATION SYSTEMS
 - D. FAMILIARITY WITH TOPIC

DIALOGUE MODE (DM)

1. ASK
2. OFFER
 - A. P. IS BEING UNDERSTOOD BY OTHER
 - B. P. INDICATES LEVEL OF DISCOURSE TOO SIMPLE
 - C. ----- APPROPRIATE
 - D. ----- TOO COMPLEX
 - E. FORMAT OF RESPONSE

PROBLEM STATE (PS)

1. ASK
2. OFFER
 - A. CURRENT ACTIVITY ON PROBLEM
 - B. PREVIOUS REFERENCE ACTIVITIES (EXTENT, NUMBER)
 - C. PREVIOUS NON-REFERENCE ACTIVITIES
 - D. CURRENT STATE OF PROBLEM MANAGEMENT

fig. 3a: Analytic rules

PROBLEM DESCRIPTION (PD)

1. ASK
2. OFFER
 - A. DESCRIPTION OF PROBLEM TOPIC (WHAT; CLARIFICATION)
 - B. DESCRIPTION OF PROBLEM TYPE - CONTEXT (WHY)
 - C. DESCRIPTION OF SUBJECT AREA (KNOWLEDGE OF SUBJECT; STRUCTURE)
 - D. TERMINOLOGY
 - E. LITERATURE

RELEVANT WORLD BUILDER (RWB) - RETRIEVAL STRATEGIES (RS)

1. ASK
2. OFFER
 - A. LOGIC FORMATION
 - B. PROBLEMS INTO SUB-PROBLEMS
 - C. PROBLEMS INTO CONCEPTS
 - D. TERMS INTO CONCEPT GROUPS
 - E. SELECT TERMS
 - F. REJECT TERMS
 - G. TERM RELATIONSHIPS
 - H. FREE TEXT VERSUS CONTROLLED
 - I. WHICH SYSTEM WILL BE USED
 - J. SEARCH STRATEGIES
 - K. FIELDS FOR SEARCHING
 - L. TERMS AS GIVEN IN EXTERNAL TOOL

RESPONSE GENERATOR (RG)

1. ASK
 2. OFFER
 - A. REQUIREMENTS ON OUTPUT (YEAR, LANGUAGE, ETC.)
 - B. NUMBER OF REFERENCES
 - C. BREADTH OF SEARCH
 - D. RELEVANCE JUDGEMENTS
-

fig. 3b: Analytic rules

EXPLANATION (EX)

1. ASK

2. OFFER

ADMINISTRATIVE EXPLANATION

A. ADMINISTRATION ON SERVICE

B. COSTS

C. ARRANGEMENTS FOR ON-LINE SESSION

D. ----- POST-SEARCH CONSULTATION

E. ----- DELIVERY OF RESPONSE

SUBSTANTIVE EXPLANATION

F. ON-LINE SYSTEMS

G. REFERENCE INTERVIEW PROCEDURES, ON-LINE SEARCH PROCESSES

H. INDEXING PRACTICES

I. THESAURAL STRUCTURE

J. DATA BASE STRUCTURE

SECONDARY COMMUNICATION (SC)

A. P. OFFERS EXPLANATION OF OWN ABILITIES

B. P. OFFERS DESCRIPTION OF OWN UNDERSTANDING OF OTHER'S PROBLEM

C. P. OFFERS DESCRIPTION OF OWN UNDERSTANDING OF OTHER

D. P. INDICATES HAS UNDERSTOOD OTHER

E. P. INDICATES HAS NOT UNDERSTOOD OTHER

F. P. INDICATES OTHER HAS UNDERSTOOD P.

G. P. INDICATES OTHER HAS MISUNDERSTOOD P.

H. P. EXPLAINS-DESCRIBES WHAT P. IS DOING

fig. 3c: Analytic rules

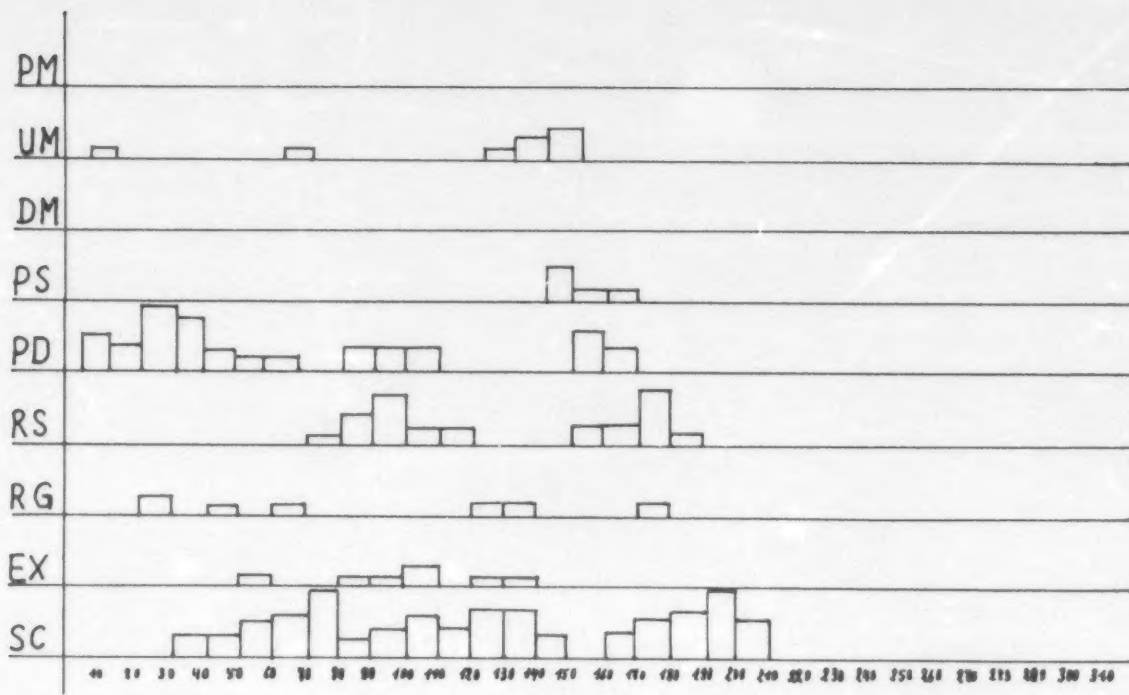


fig. 4: Interview 4, User Functions, by 10.utterance groups

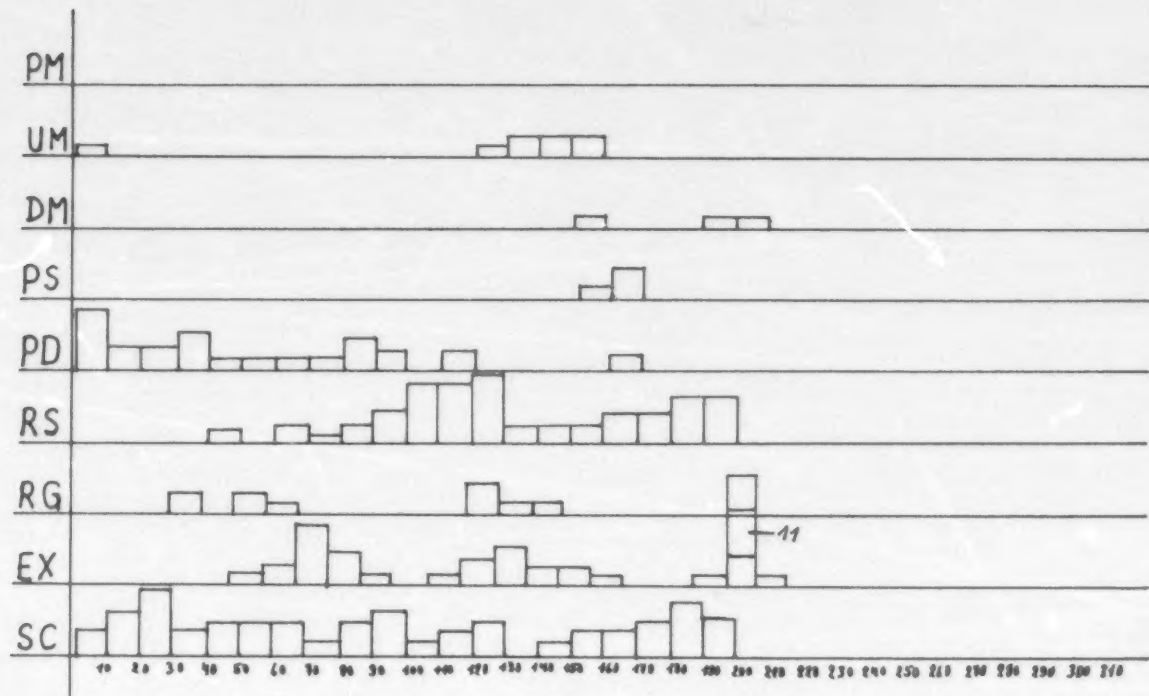


fig. 5: Interview 4, Intermediary Functions by 10. Utterance groups

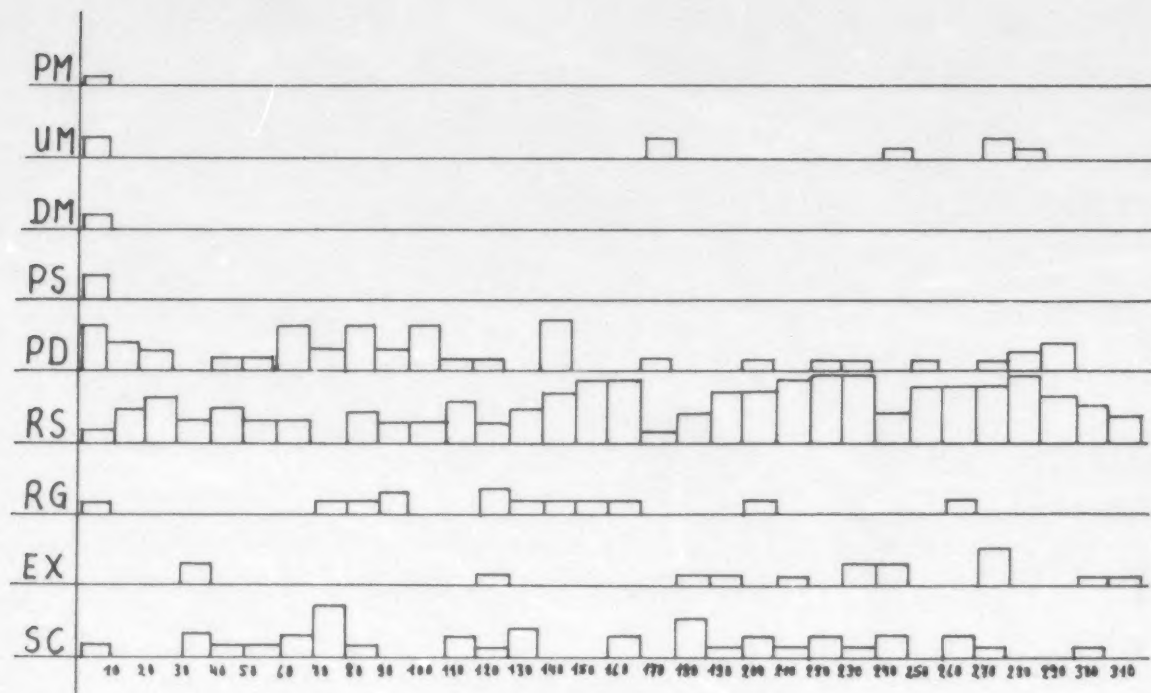


fig. 6: Interview 5. User Functions by 10-utterance groups

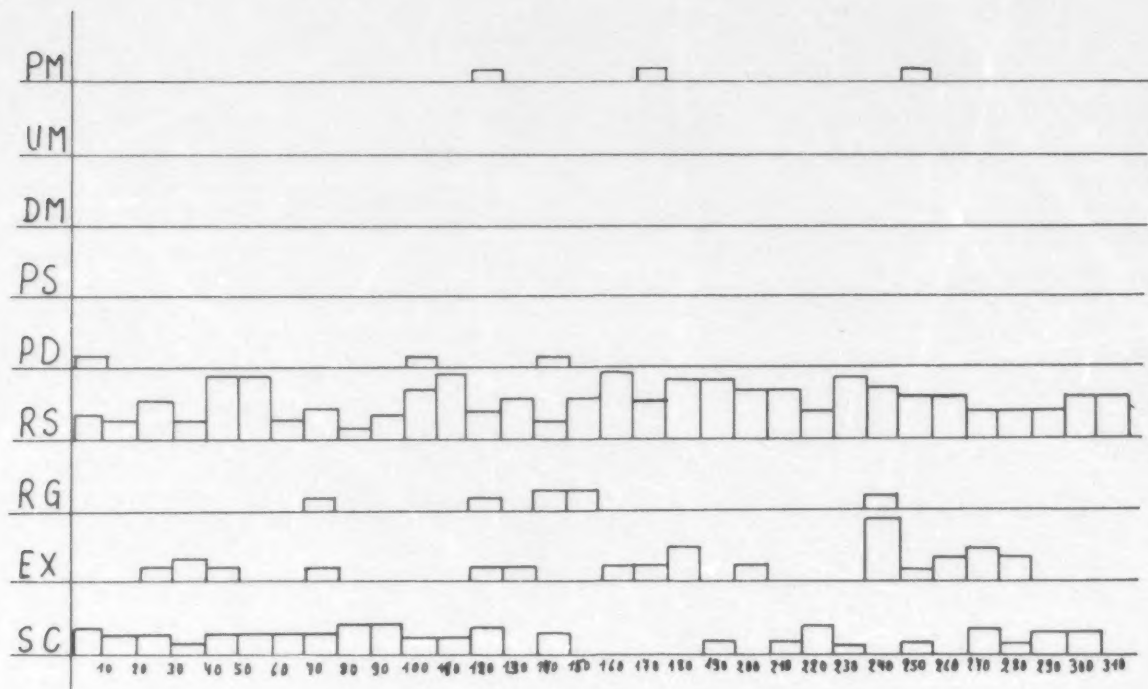


fig. 7: Interview 5, Intermediary Functions, by 10. utterance groups

4. CONCLUSIONS

4.1 Interpersonal interaction

If we take the MONSTRAT model as a general frame of reference for analysing and evaluating the intermediary's role in online retrieval processes, as represented by these data, we have to draw some conclusions which seem to be far from being optimistic.

Substantial help for the user seems to be possible in only two narrow rather limited phases of the whole process.

First, it is the starting phase, where the intermediary may lead the user to get a general overview which may result for example, in a concretisation of his problem.

Second, in the final phase of the problem treatment process, where, for example a confirmation of the users strategy of action is needed, the intermediary can be helpful by providing the respective output of references (and abstracts) which support the strategy.

During all other intervening phases the potentials of the intermediary (and system) are rather limited. The whole process remains a static one, dynamic courses of interaction which relate to the respective phases of problem solving do not take place. To summarize, intermediary (and system) do fulfill a real help function only in those phases, where the provision of knowledge for problem treatment does not play an overriding role. As a result, the inherent incompetence of the system often forces the intermediary to use suboptimal and counterproductive communication strategies. Dominated by the system's restricted capabilities, i.e. the reduction of all components to a one dimensional form (references), the intermediary, without getting substantial support is more or less forced to retreat to a position where technical professionalism, but not actual help for problem treatment, plays the major role.

4.2 Problem description

From these data, it is clear that it is possible to isolate occurrences of the problem description function in the advisory interaction, to identify specific tasks performed within problem description by the advisor, and to demonstrate that problem description appears to be necessary for successful information interaction. Furthermore, it appears that in performing this function, advisors very quickly decide upon a stereotypical problem type structure, which subsequently drives not only the problem description function, but many of the other MONSTRAT functions as well. Indeed, this type of structure appears to offer a means for integrating performance of the various functions. And, finally, we note that in the advisory situations, unless the problem were of a particularly highly defined sort, interaction between user and advisor was necessary in order to describe the problem sufficiently for advice to be given.

4.3 On-line functions

These results lead us to conclude that, in certain circumstances, including experience of the intermediary in acting as one, most of the MONSTRAT functions can appear in the on-line pre-search interview, and that, in fact, some of them appear to be necessary for successful interaction.

It is instructive to compare these results with those of the 'interpersonal interaction' investigation, since they seem at some points to contradict one another. In interview 4, for instance, the intermediary performed the problem description function a great deal, attempting to understand the user's problem and to help the user to clarify the problem sufficiently for effective search. In interview 5, however, the intermediary concentrated almost exclusively upon the RWB function, in much the same way that the intermediaries of the 'interpersonal interaction' investigation did. There were significant differences between intermediaries 4 and 5 in terms of experiences, 4 being much more experienced, and we conjecture furthermore that the social and speech situation of interview 4 & 5 were much different from those of the 'interpersonal interaction' group. In the latter, for a number of reasons, the intermediaries perceived their major problem, and hence role, to be deal-

ing with the vagaries of the on-line system itself, hence the almost total lack of "real" interaction with the user. These results lead us to conclude that, in the absence of constraining factors, intermediaries will perform the MONSTRAT functions in a highly interactive way with users, and that such interaction provides the best chances of success in the search.

General

The results of these three largely independent empirical investigations have thus demonstrated that:

it is possible to apply the MONSTRAT model to information interaction as a analytic tool;
the MONSTRAT functions do appear in a variety of information interactions, and can help to show common features among these types;
analysis of information interaction can suggest how IPM performance could be improved or how better IPMs might be designed; and especially
that interaction is absolutely crucial to successful performance of at least some of the MONSTRAT functions.

References

- (1) Progris: Die Rolle des "Intermediary" in on line-Informationsprozessen - Eine Exploration -. im Auftrag von Projekt INSTRAT. Berlin, Freie Universität, FUB-IFB 11/82, 1982.
- (2) Belkin, N.J., et.al.: Description of the MONSTRAT problem description function. Projekt INSTRAT, in preparation.
- (3) Brooks, H.M.; Belkin, N.J.: Using discourse analysis for the design of information retrieval interaction mechanisms. ACM SIGIR International Conference on Research & Development in Information Retrieval. ACM, Washington, D.C. 1983.
- (4) N. Belkin, T. Seeger, G. Wersig: Distributed Expert Problem Treatment as a Model for Information System Analysis and Design. In: Journal of Information Science, 5(1983), S. 153-167.

Two-dimensional Representation of Information Retrieval Services

H.P. Frei

Winterthur Life Insurance Company
Winterthur, Switzerland

J.-F. Jaussin

Swiss Federal Institute of Technology (ETH)
Zurich, Switzerland

Two-dimensional Representation of Information Retrieval Services

H.P. Frei
Winterthur Life Insurance Company
Winterthur, Switzerland

J.-F. Jaustin
Swiss Federal Institute of Technology (ETH)
Zurich, Switzerland

The user-machine interface of an information retrieval system is presented. This interface supports browsing through information structures and information items as well as formulating and processing queries. Both information and commands are structured either hierarchically or in a network-like fashion. These structures are displayed in a two-dimensional manner on a graphics screen and can be explored by means of motion commands. Queries are formulated by filling in the fields of a template.

The system is implemented on a relatively small personal computer and the software is written in a high-level system programming language.

1. Introduction

While nuclear power, space exploration and the energy crisis were some of the important topics of past decades, the key issue of today's post-industrial society seems to be information. At the present time information plays a role similar to the role energy used to play. It is one of the most valuable resources for industry, organizations, and individuals. However, the outstanding significance of information is not a product of recent days. For a long time information has been both a costly merchandise and a substantial element of power. In recent years, novel technologies changed the handling of information significantly. As a consequence, supply of and demand for information is increasing considerably nowadays.

Information is an important base for every problem solving process, and for many activities it is to be considered a resource similar to energy, finance, equipment etc. We are used to carefully managing these resources and we have to learn to manage information in a similar way. As the amount of information to be handled is usually rather large, the usage of modern technology for organizing, storing, processing, and retrieving information is a must.

Various institutions are developing novel *Information Retrieval (IR) Systems* and an important characteristic of these systems must be their universal applicability and extreme user orientation. The *user-machine interface* of such a system is the topic of this paper. It is a topic of utmost importance as not only the amount of machine-readable data is growing but also the diversity of computer users. It is to be noted that the majority of computer users in general no longer consists of EDP specialists. Most of today's users are employing computerized systems only occasionally. In addition, *IR Systems* are increasingly being used by *casual users* in the course of various kinds of problem solving processes.

2. The Caliban System

The user-machine interface reported here is the one of an *IR system* called *Caliban*, developed at ETH. The main objective when designing *Caliban* was to have a small and inexpensive but nevertheless user oriented *IR system*. The system was to support *formal retrieval* by means of queries as well as *browsing* through both information structures and information items. An important focus was on exploiting the human being's highly developed spatial memory and ability to recognize graphic cues. In addition, we tried to use the instinctive human skills for large parts of the command dialogue.

Caliban does not include routines to automatically analyse information; nor did we try to build up our own information structures. Instead, *Caliban* was designed to be as independent as possible of the particular kind of information to be handled. We refrained from putting together our own small test collection; rather, we tested the system with commercially available bibliographic data. (This test data

was purchased from INSPEC in machine readable form and is on physics, electronics, and control.) In addition to the bibliographic data we bought the corresponding *information structures*, i.e. a classification and a thesaurus. Choosing commercially available data rather than an experimental data set was particularly challenging. We encountered a number of properties one would hardly have thought of when collecting test data.

Caliban is implemented on the personal computer *Lilith* [W1 81], with a raster scan display of 594*768 dots as output device. Input is given via a standard keyboard and --most important-- via a mouse pointing device. In addition, the three buttons of the mouse are used to invoke commands as will be elaborated later. The software of *Caliban* is written in *Modula-2* [W1 82], a Pascal-like high level system programming language which allows separate compilation of modules and provides accurate type checking.

3. Displaying Information Structures

Humans often tend to structure information in a hierarchical manner. A well known hierarchical partitioning of knowledge is the decimal classification. Genealogical trees or a table of contents in a book are other typical examples of this kind of representation. It seems to be natural for human beings to split general concepts into more detailed sub-concepts. Furthermore, hierarchical representations are used widely enough to allow someone, who did not define a particular structure, to easily understand its meaning.

Unfortunately, the nature of information is sometimes more complex than a pure hierarchy. A network is often the appropriate structure. A typical example is a thesaurus where there are a number of connections (e.g. to related terms) which do not fit into the hierarchy. *Caliban* is able to handle both hierarchies and networks.

As many information structures (e.g. classifications, thesauri) are of significant size, the problem arises how to present them to the user of a system. There are two limiting factors. First of all, the amount of information a user is able to perceive is restricted. Secondly, the dimension of a screen is rather limited and only parts of a sizable structure can be displayed; in addition, if a graphical presentation is chosen, many of the edges of a structure would lead to nodes outside the screen. This applies to both hierarchies and networks. However, edges of a hierarchy crossing the boundaries of a screen invariably lead to a node of the immediate structural neighborhood: a father or a son. Contrarily, edges of a network crossing the boundary of a screen may lead to an arbitrary--e.g. structurally far distant--node of the network. As we felt that such edges would confuse the user, we decided not to display them on the screen.

We ended up capturing the important relationships of a network as a hierarchy. Only the hierarchy is presented to the user in a graphical manner despite the fact that there are further connections in many instances of information structures. In this way, *graphically* represented trees are displayed to the user instead of networks. However, the user may request additional relationships in a *non-graphical* representation.

Because of the restricted size of the window in which we display information, only three levels of a tree can be graphically displayed. Nodes are represented as rectangles with their names written inside and the father-to-son-edges of the structure are explicitly drawn as line segments.

The node in the center is called the *current node*; it is highlighted to distinguish it from the others. Only the father of the current node, two brothers and a relatively small number of sons are displayed at the same time. This view of the tree can be changed in two different ways: either a new current node is chosen--which will then appear in the center of the window--or invisible sons of the actual current node are made visible. Dots appearing above the leftmost or the rightmost son of the visible part signify that more nodes are available in that direction.

In our test data one of the index structures is a thesaurus. It is a typical example of a network with a hierarchical main structure and a number of other connections such as connections to broader terms, related terms, top terms, and classification codes. The broader terms represent more general concepts in addition to the father term which exists all the time. Related terms are terms with a direct relation to a particular node term other than the nodes directly connected through the tree structure. Top terms are the possible ancestors and classification codes represent connections to the outside of the index structure, namely to a classification tree.

In contrast to the pictorial representation of the hierarchical part of the structure, we chose to represent these additional connections in a textual manner.

4. Services and Their Representation

Although many of the services of an interactive system are conceptually divided into several sub-services, many systems present the sub-service commands on the same level as the service commands. The casual user encounters the following problems: If a system is powerful, there are a great deal of commands available and it takes a long time to memorize a particular command subset for using the system in a simple way; as many of the command sets consist of linear command lists, it is difficult to find out in which state the system is, i.e. which commands are meaningful and/or executable. If the services of an interactive system are structured, another problem arises when a command is given: the system generally enters a new mode, i.e. the previously active environment is left. The set of possible commands changes which either means that new command names become available or old ones change their meaning. The difficulties for the user mainly originate due to the lack of information provided by the system. It is our belief that, if the services of a system are structured, this structure should appear on the screen.

The *Caliban* services are structured and--as we had already a tree representation for the data--we decided to use the same graphical way to represent our services. This is in contrast to referring to a linear list of commands and has a number of advantages. A service is a part of the system and constitutes an exactly defined activity, a *state of mind* of the user. Therefore, only those sub-services and commands relevant to the particular service have to be active and executable. As soon as another service is entered, other sub-services and commands become relevant and active.

There are two ways of invoking commands in *Caliban*. The first one is to point to a specific position on the screen, say a tree node or a so-called soft key. This quick and powerful way to invoke commands is accomplished by first putting the cursor at the desired position and then pressing any of the buttons of the pointing device. In order to eliminate any kind of doubt, the system always illuminates the object when the cursor is placed on it. In this way, the user knows exactly in which state the system is and to where he or she is pointing. The second way of executing commands is by pressing a hard key either on the keyboard or on the pointing device (mouse) independent of the position of the cursor.

In *Caliban*, pressing hard keys on the keyboard is reserved for commands which have irremediable consequences, say deleting an item. Pressing a mouse button is a very fast way of invoking a command, especially when there is no need to position the cursor to a specific location. This mode of execution is reserved for commands having minor consequences if they are entered by mistake. Most of the time the combination of positioning the cursor and pressing a mouse button is used. As different meanings can be assigned to the buttons of the mouse when pointing to different locations of the screen, an explanation of the three buttons is always given to the user in the *Message* window, which is a useful help for the casual user.

At the outset, after the system has been started, all the three buttons of the mouse have the same meaning as shown in Figure 1: they serve to *EXECUTE* a command. The cursor has the shape of a mouse and is visible in the *Command* window near the node *Browse* (Fig. 1).



Fig. 1 : Situation after starting Caliban

Caliban provides two different ways of getting information :

- *Browse*, meaning to move about index structures similar to inspecting the books on the shelves of an open library.
- *Search* to prepare a query and submit it to the system in order to retrieve items.

Browse and *Search* are the two principal sub-services of the Caliban system. The two other sub-services visible in the *Command* window (Fig. 1), *Edit Set* and *Combine Set*, allow the combination

of retrieved sets of items.

5. Browsing

One of the main objectives of the system is to support and even suggest associations of thoughts to the user and to supply a means of following the different paths opened. This is particularly important when the user has only a vague idea of what he is looking for. In this case the user needs a browsing facility which allows to collect terms that can be used in a query formulation. The graphical representation of the index structures on the screen facilitates finding ideas for future use. In addition, the user gets an impression of structure when explicitly moving to the father or to one of the sons of the current term or when inspecting further relationships. This kind of browsing through information supports an associative way of thinking by humans.

Browsing also allows to inspect real information items attached to the index terms. This gives a more precise idea of the meaning of a particular term and is similar to what the patron of a small library can do by opening a few books when inspecting the shelves. Furthermore, we are convinced that in many cases a convenient browsing facility serves to directly find the requested documents without formulating a formal query.

As mentioned in 2, we employ bibliographic test data purchased from INSPEC. Among the five index structures *Caliban* provides, there is a classification and a thesaurus. Every node symbolizes an index term and there is always a set of documents attached to it. In contrast to the classification, which is a regular hierarchical structure containing about 2,600 nodes, the structure of the thesaurus is a network containing about 6,200 nodes. In addition to these two index structures purchased from INSPEC, we built three others: free-indexing-terms, author-editor, and publisher. They simply consist of trees with nodes that are alphabetically ordered on each level. An intermediate level of indices was introduced where deemed necessary.

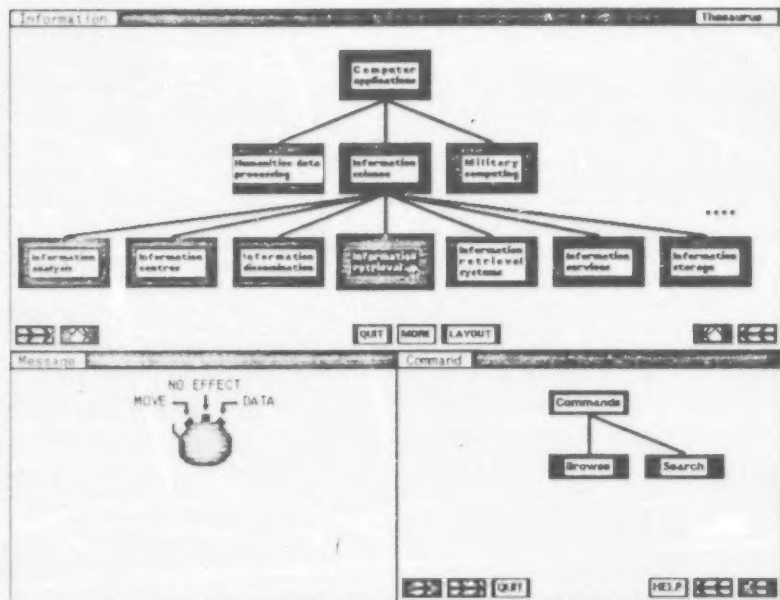


Fig. 2: Tree representation of an index structure

Figure 2 shows a typical layout of the screen when browsing. The user has first moved to the node *Computer applications*. He then decided to choose *Information science* as current node and now points to the term *Information retrieval*. If he were to press the left button of the mouse (MOVE) while pointing to that node, the term *Information retrieval* would become the current node and a new environment would appear on the screen. By pressing the right button of the mouse (DATA), a set of items connected to the node pointed to will be displayed on the screen.

The dots above the rightmost term (*Information storage*) mean that there are more than seven sons attached to the term *Information science*. Four of the soft keys at the bottom of the *Information*

window can be used to make invisible sons visible. More specifically, in our example the double arrow in the lower right hand corner of the window allows to shift the node *information storage* to the leftmost position on the screen, thus displaying six more sons.

If the index structure is a network, as in the case of our thesaurus, the user can request more structural information about the current node (*MORE* soft key). More --not hierarchically related terms-- will appear on the screen in a non-graphical way. If one of these terms seems to be of interest, the user can *jump* to appropriate node which will become the current node. The result is a similar representation of terms in Fig. 2.

As mentioned earlier, the IR system *Caliban* has been designed to treat various kinds of information such as office information, letters, memos, technical reports, personal data, etc. We demonstrated the capabilities of the system using our current test data, bibliographic material. Although other data would certainly be structured in a different way and therefore represented differently, the browsing facility on the corresponding index structures would be the same.

Browsing is to be distinguished from other forms of retrieving information :

- Retrieval exclusively depends on the particular decisions made by the user while moving about the index data; however, the user may follow his or her associations of thoughts which is, in many cases, a concept superior to asking plain questions. On the other hand, a great deal of information can be missed when the retrieval depends on more or less random paths.
- The browsing process does not allow one to access data from several parts of index structures at the same time; retrieval is limited to asking for data connected to one term only. The retrieved items are linked only by that term.

The next chapter explains a completely different approach to retrieving information items.

6. Query Formulation

When formulating a query, the user has to fill in the fields of a query-template rather than using a command language. In Fig. 3 the user is preparing what we call a *virtual information item*, similar to filling in a row of a sample relation in query by example [ZL 75]. Each searchable field of the stored information items, such as author, publisher, classification, thesaurus or free indexing terms in our test data, may be left blank or filled in according to the information the user is looking for. If a user is looking for items on a specific subject written by a specific author, the corresponding fields have to be filled in. Then, the virtual information item is submitted as query to the system which will return a set of real information items corresponding to the virtual one.

However the user does not need to type long and complicated sequences of characters: the tree-structured indices are always available (SHOW INDEX TREE) to assist the user when preparing a virtual information item. Terms may be picked directly from the screen and included into the virtual item.

The MOVE CURSOR command (Fig. 3) allows to explicitly select a field of the template in order to fill it with query information. No matter to which field the cursor points, the index structure corresponding to that particular topic may be displayed by invoking the command SHOW INDEX TREE. Although in search mode, the user may use all the BROWSE facilities described above, browsing becomes a sub-activity of searching (Whether browsing should also formally be treated as a sub-activity of searching and appear as a brother node of Find, Modify, etc. in the command tree is debatable). The important point is that interesting terms can automatically be included in the query formulation, i.e., into the virtual information item. In fact an arbitrary number of terms accessible through MOVE commands can be selected and included in the virtual item. Choosing is accomplished by pressing the

middle button of the mouse while pointing to a node. In contrast to Figure 2 where the user was employing the browse service, the meaning of the middle mouse button is now CHOOSE (instead of NO EFFECT) when the user is browsing in the search service. That is the only difference from the situation of Figure 2.

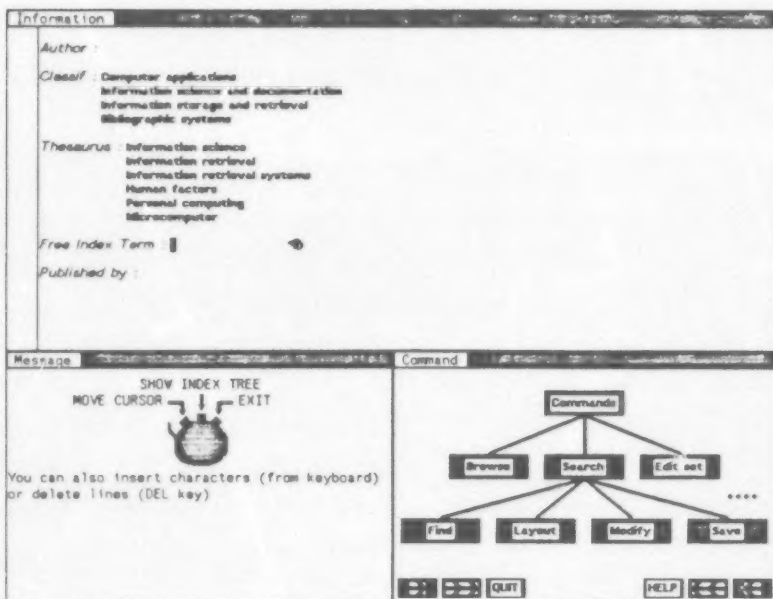


Fig. 3: The query template with some terms

The index tree is left by the QUIT command and all the items chosen appear in the template. The user could now select any other index structure (or even the same again) in order to CHOOSE more index terms. In Figure 3, he or she selected first the classification tree and picked four terms, namely *Computer applications*, *Information science and documentation*, *Information storage and retrieval* and

Bibliographic systems. Six other terms have been taken from the thesaurus.

Once the template is satisfactorily filled with query information, the command *Find* is activated to retrieve real information items. Prior to executing the actual query algorithms, the user gets the opportunity to weight every field according to the importance a term should play in the retrieval process. These weights are used by both the retrieval and the ranking algorithms. The retrieval algorithm retrieves those information items which exhibit a higher similarity to the virtual information item than a certain threshold. The ranking algorithm ranks the set of retrieved items by assigning higher ranks to items with higher similarity to the query.

After a set of real information items has been displayed to the user it is possible to choose an arbitrary item as the next query. This is made possible by the fact that the structure of a *Caliban* query is identical to the structure of a real information item. In addition, the user may select any terms found in any of the items and include them in the template initially prepared. Other modifications like including new keywords or deleting old ones, can improve the quality of the search. In this way, the quality of retrieved item sets can continuously be upgraded in an *iterative process*. The idea of directly taking into account results of previous searches has already been reported by Oddy [OD 77] who describes a system which enables one to insist on terms belonging to a retrieved item or remove those terms from the query which seem to be outside the domain of the search.

There are two major advantages of such a retrieval process: first of all, a query may be prepared by directly employing index structures; secondly, the query formulation is very *similar* to the information item.

7. CONCLUSION

In order to make flexible information retrieval possible, powerful interactive systems are necessary. The IR system *Caliban* has been designed and built employing those modern hardware and software

tools emerging at the present time. As far as the hardware is concerned, the user friendly input and output devices are especially noteworthy. Software-wise the advent of high-level system programming languages facilitates the development of software a great deal.

However, one of the principal aims of information retrieval research must be to remove the still existing barrier between the persons seeking and the systems containing information. Developing the Caliban system was a step in this direction. Easy-to-understand concepts underlie Caliban's user-interface. These concepts include the notion of browsing and the notion of the virtual information item which seems to be a natural way of expressing a user's information needs.

References

- [BA 83] Bartschi, M., and Frei, H.P.: Adapting a Data Organization to the Structure of Stored Information. *Lecture Notes in Computer Science 146*, Springer-Verlag, Berlin, 1983, pp. 62-79.
- [FR 83a] Frei, H.P., and Jauslin, J.-F.: Graphical Presentation of Information and Services: A User-Oriented Interface. *Information Technology: Research and Development*, Vol. 2, 1983, pp. 23-42.
- [OD 77] Oddy, R.N.: Information retrieval through man-machine dialogue. *Journal of Documentation* 33, 1977, pp.1-14.
- [WI 81] Wirth, N.: A personal computer for the software engineer. *Proc. of the 5th Int. Conf. on Software Engineering*, San Diego, 1981, pp. 2-12.
- [WI 82] Wirth, N.: *Programming in Module-2*. Springer-Verlag, Berlin, Heidelberg, 1982.
- [ZL 75] Zloof, M.M.: Query by example. *Proceeding of the NCC, AFIPS*, Montvale, N.J., 1975 pp. 431-438.

Irving Louis Horowitz (Rutgers University, New Brunswick, New Jersey, USA)

"New Technology, Scientific Information, and Choices for Democratic Societies"

This study examines the implications of information technology for the political prospects of democratic societies. This subject has been largely overlooked in the literature of communication and the social sciences. The paper will address this problem from the standpoint of political sociology. It divides attitudes to information technology into Orwellian (fearing the power of information technology over the masses) and Marxian (fearing a threat to public ownership or popular expression). It specifically discusses the effect of information technology on the publishing industry; but also discusses the long-term political effects of the new technology in other media frameworks.

INTRODUCTION

A broad range of literature is available concerning the worth of the new information technology at the level of business and industry. This new information technology has become sufficiently large to merit its own slogans, the more recent being "the videotex' revolution."¹ However, the implication of this new technology for the prospects of democratic societies have largely been overlooked. The author addresses this question of the political impact of information technology, not from the vantage point of contending commercial proprietary interests or to make technical distinctions between delivery systems, but rather, primarily, from the standpoint of sociology. More exactly, what are the effects of information technology of societies and citizens as a result of developments that come under a variety of labels; from electronic data to information technology, broadcast teletext to interactive videotex; but which add up to the same thing--the dissemination of information from senders to receivers in formats without benefit of (or supplemental to) hard copy, i.e., the printed word.

Responses to the new information technology seem to fall into two camps. The first is grounded in what might be called the Orwellian or counter-Utopian model. Information technology is feared as inevitably dominated or controlled by a small political clique or power elite, capable of maneuvering and manipulating mass sentiments. In this respect, there have been dire predictions that the new information technology is some awful majesty, a veritable handmaiden of totalitarianism. An entire tradition has, like Orwell himself, come to view with deep concern any control or domination of the masses by electronic media as such.² Critiques of the totalitarian potential of the new information technology extend from an established position taken in English scholarly circles from Aldous Huxley to C.P. Snow, no less than American social and cultural fears from Dwight MacDonald to

Marshall McLuhan. The key issue was not, and is not, about the quality of the hardware or the data being transmitted; but rather ownership and control. Information technology in this context becomes only a larger and more ominous evolution of a long-standing process of centralization of authority.

At the other end of the political spectrum are those who have been critical of information technology as a "wholly other" culmination of the imperial capitalist state through monopolization of the means of scientific communication.³ Marxists see in the new information technology a definite threat to potential for public ownership or popular expression, usually only in Western societies. There is a fear of the new, a mistrust of the potential for the management and manipulation of ideas, not so much by a political elite as by an economic ruling class. If one looks at the two main lines of criticism, their conclusions and arguments present striking parallels.

The question for both Orwellians and Marxists is essentially the same: who will dominate the political order? Their bitter differences lie more with which elites are identified as the main enemy of freedom, rather than attitudes toward the significance of new technology as such. In this regard, both those who fear the totalitarian implications and those who fear class domination of the new information technology are very much part of a long-standing difficulty in understanding an advanced society in transition. There is serious doubt that a society in which traditional modes of expression become obsolete can at the same time hold firm to traditional forms of political rights. Although the extremes represented by these antipodal positions fuse, the analytical problem remains the same. Those who fear totalitarianism and those who loathe class domination share an assumption that no positive political goods will evolve from the new information technology.

LONG TERM EFFECTS OF INFORMATION TECHNOLOGY AND DEMOCRACY

The concerns that have been expressed in the publishing establishment,

which will be most directly affected by a shift from paper to videotext formats, are specific extensions of the general premonitions and predispositions shared by political leaders of developed nations. While such concerns are often couched in the rhetoric of business prospects and profits, they are no more or less the kinds of concerns which have occupied literary and scientific people working in the areas of mass communications throughout the Twentieth Century, down to the hushed atmosphere of impending doom. There is nothing in the evidence to suggest that the flow of print has contracted under the impact of the new technology. As one report indicates:

In 1950, when the impact of television first began to be felt, 11,022 books were published in the United States. In 1970, when the impact of the computer began to reach major proportions, the number of books had risen to 36,071. In 1979, after almost 30 years of television and 10 years of major computer use, 45,182 books were published in the United States. Book publishing revenues in the United States in 1950 were less than \$500 thousand; in 1970 they were more than \$2,900 thousand; in 1980, more than \$7 million.⁴

Even with a levelling off, the multi-tiering of information is fact, not conjecture.

This deep background indicates the existence of democratic potential in the new information technology. By no stretch of the imagination will these potentials be automatically realized. However, the notion of democratic potential is worth pursuing in at least five components:

- the increased amount of information available in the new technology
- the necessarily active role of the participant in much of this new technology
- the corresponding capacity for confirmation and verification of information to an extent previously unavailable
- the public rather than privileged nature of this new information technology
- a redefinition of the very structure of information and cultural climates within our society.

It has been argued that the sheer abundance of information creates the basis for a system overloaded with such "information pollution." There is supposed

to be a corresponding acceleration of "narcotizing dysfunctions" of mass media. While this "drugging" of society has in the past been the most widely identified concern of researchers, the point has too often been made of the cost of any appreciation that the massification of information is by no means an intrinsic evil.⁵ So much of the post Second World War sociological literature on information and communications was laden with moral imputations about the risks and dangers of succumbing to the blandishments of "bureaucratic" or "mass" society, that the explosion of possibilities opened up by radio, television and satellite communications became strangely converted into threats and predicaments for intellectuals.⁶ Even those who dared adopt a more level-headed and realistic appraisal, tended to channel their thinking about information technology into questions of leisure and mass culture.⁷ One has to search the social science literature far and wide to come up with a serious analysis of media and politics or information transmission and democracy.

The discussions which are now beginning to take place on the relationship between information technology and democracy are not unlike the earlier literature described above. There is the same tendency to perceive of recent developments in science as favoring elites, dooming masses, narcotizing recipients, manipulating the poor and massaging the wealthy. But if we keep that earlier literature in mind, and recognize quite frankly that its most dire predictions have not come about, or at least have been counterbalanced by serendipitous aspects of communications, we can then move to higher ground in a discussion of the new information technology as a set of possibilities concerning democracy and free expression, rather than a predetermined unilinear totalitarian outcome.

A modular, grid-like pattern of information storage and retrievability such as that permitted by videotext, also allows for a wider level of choice

and decision than previous forms of telecommunications. If one takes into account that standard television viewing provides for a maximum of 99 channels to be used as separate outlets on two main frequency outlets, and if we multiply that by 99 additional outlets within each of the original channel separations, the outcome is no less than 9,801 possibilities at any given instant of video viewing. That is to say that there are roughly 10,000 active possibilities available through broadcast teletext at any given time.

These may range from shopping market lists, updated information on industrial patents, air schedules, to musical concerts and experimental dance. New forms of interactive communications, such as videotext, provide even higher levels of retrieval, storage and utilization of data to wide numbers of people who in the past have been largely recipients of such information.

Despite the fear that interactive videotex may mean the end of the printed word, the more likely proximate consequences are less portentous. We are clearly moving toward a multitiered system: one based on both hard copy and videotex (or videodisc) information bases. They serve very different but mutually important purposes. A parallel would be radio and television: television did not replace radio, but it did influence the content disseminated by the older medium. With videotex, we are now at a point where each family, individual or business unit may have at their disposal 10,000 different "titles" or forms of upgraded information (far more than any average home or office can presently maintain, much less store) and, in addition, hard copy for more abstracted and more portable forms of information. But the act of broadening the amount of information accessible through technological devices opens up channels of choice and decisionmaking for individuals that hitherto have only been available to powerful elites within a society.

We are already living in an environment which has a multilayered media

media system. The audience which continues to buy a mystery novel or a philosophical treatise in hard copy form, will also insist upon having a personal computer with which to access financial databanks and update their portfolios. The individualization of information and not the collectivization of ideology is what is realistically beginning to occur. Policies to implement an information environment which remains competitive thus become a democratic imperative. The key is competition and not monopolization of the sources of ideas. The nature and history of scholarly publishing uniquely equips this dissemination center to act upon such concerns.

PERSONAL INVOLVEMENT AND PUBLIC TRUST

The new information technology encourages active rather than passive public involvement. Whatever virtues are possessed by traditional printed material (literality extends from ease of use and retrieval to shifting back and forth within a journal or book), the act of reading a given book or periodical remains relatively passive and singular. How information is structured, and moreover, what is not revealed, is decided by others. Unless one is dogged and determined, it is difficult to verify data or check possible sources of confusion, misunderstanding or even downright error outside of library compounds. Alternative information sources still go largely unchecked.

The new information technology permits a higher level of interactive involvement. One can confirm or disconfirm exact information, test propositions and develop comparabilities not foreseen by the author or the original source in the comfort of one's home. In discussing confirmation, we are generally considering factual rather than interpretive information. Even at the level of speculation, there may be possible uses of information technology that were hardly considered in the past, e.g. tracking sources of ideas through space, time and culture. The paired acts of verification and

confirmation may become part of the everyday life of ordinary individuals, rather than an exceptional event requiring extraordinary effort and skills by specialized elites. Optimally, in the event of further data specialization, new forms of generalization become possible.

In interactive videotex systems, we may be moving beyond a narcotizing and dysfunctional environment in which the individual passively receives images and signals, as in broadcast teletext. The multiplication of images may lead to higher levels of narcotizing behavior, but it is more likely that this will stimulate a computer involvement which will permit not only verification and correction of information but also the capacity to relay information from the individual back to various stations or databanks with which that person is networked. Voting behavior, for illustration purposes, may change dramatically, as it becomes possible to cast ballots through a hookup between a decoder-equipped television receiver and a main system computer. Opinions and votes may be cast on matters of local and parochial interest no less than national interest. One can receive and view the activities of local boards of directors of corporations, local courtrooms, federal hearings or government political offices. Optimally, this will be plugged into a cable access system permitting the individual to see, hear and register his personal points of view and persuasions almost instantaneously. What is being developed is not simply heightened technical complexity, but a decline in human passivity, and hence an enlargement in the scope of the democratic process. This is a process linking not simply individuals, but communities and continents as well.

The importance of the publishing process will not necessarily diminish, since with such a massive set of availabilities made possible by electronics, the need to select, screen and simply search out what is relevant itself becomes a special province of the publisher.⁸ The act of making information

public has always been closely identified with the process of marketing. One must surmise that in the new information technology environment, marketing activities will both increase and become more focused. This is especially the case in the short run. High start-up costs in computer information devices mean that they will be more profitable and competitive in a "narrow" rather than "wide" information band. For example, the capacity to retrieve information electronically expands rather than retracts the amount of hard-copy requirements. For however active recipients of information may become, the need for intermediary layers of sifting data and ideas will not disappear, but in all probability will expand.

The tendency to exaggerate technological innovation, so that its apocalyptic features are exaggerated while the cumulative features are negated, has taken on dangerous proportions. The prophets of the new information technology, in their emphasis on a transition from a paper-based to a paperless society, are invoking a mystical trend. The idea has taken root that advocates of a continuing role for books and journals are somehow offering obstacles to progress.⁹ Arguments have been presented in terms of the resistance of the academic community to view publications in electronic form as legitimate: the inadequacy of copyright in dealing with a fluid situation in which text can be transferred from creator to user, loss of revenue from advertising sources etc. But these fears often rest on the notion that a single-tiered, paperless world is an inevitable outcome of the new technology. Some researchers have come to take for granted the impending paperless society, and are already turning their attention to the social consequences of a world without books.

But before sounding such alarms, we should observe the obvious anomaly that such dire predictions related to the demise of the printed word are invariably registered in conventional books and journals. The larger problem

in the single-tiered approach is a failure to reckon with great variety in the structure of thought. Organized rationalized conceptions are of three sorts: information, ideas and interpretations. While one can readily appreciate how legal documents, government economic statistics and specialized abstracting services can be designed to displace conventional journals and reference volumes, the need for ideas, and even more important, the play of ideas that permits a range of interpretations, are not readily subject to a paperless environment. The most potent argument against the anti-Utopian vision is that a government's frightening power to control life and crush opposition in an abstract sense is, in fact, restricted by the greater availability of multiple forms of information dissemination, idea construction and interpretative analysis.

The history of science and technology indicates that the latest and newest modes of communication and transportation do not liquidate the need for earlier forms, but become a value-added phenomenon. While supersonic jet travel may make sense over great global distances, conventional jets at subsonic speeds may work more economically over shorter national routes. Indeed, for distances under five hundred miles, railroad and boat travel may suffice, and may even be best. Beyond that, surface transportation like automobiles, buses and trucks are clearly best for local travel. One of the most disastrous events in U.S. transportation history occurred with efforts to replace the railroad with the aeroplane. This only created a crisis point at which the costs of convenience of transportation became exorbitant. Railroad travel has had to be restored and stations vastly improved and even new ones build. There is much in this slice of transportation history that should inform the communications industry. Once again, we are faced with a modest add-on view of new technology confronting an apocalyptic view that such a technology does away with the need to transmit ideas and

and interpretations (and even information) in paper form. Those for whom democracy is a central issue, would do well to recognize that the value of a multitiered approach is a question of ideology no less than empirics.

Before publishers worry about big brother controlling the big tube, it might be wiser to resist the temptation of exaggerated premises about scientific manipulation, and to seek to utilize new media for the further promulgation of technical and scholarly materials.

It has long been held that the television screen has helped privatize individual life in the United States; that its widespread use led to a decline in public participation and an increased willingness to receive images without living a life. The extent of such consequences remain moot. But we may reverse privatization given the magnitude of the new information technology. The signals that individuals are able to send by cable or telephone may make for enlarged public and civic participation. It is probably true that such public roles will appear as street manifestations or public outcries. But on the other hand, the scope of interactive videotex participation is such as to add, rather than subtract, meaning to the notion of civic participation. This will become increasingly clear as information technology becomes more widely available. It may not be possible for everyone to be constantly involved in every phase of decisionmaking, nor would citizens wish to do so.

The expansion of democracy does not entail any diminution in specialization. In its offer of abundance, the new information technology implies a check on, as well as a reliance upon the specialist. Just as academic life itself has become increasingly specialized, so too will these new developments in information technology, on a consumer level, move towards the servicing of a highly refined and limited clientele with particular interests and tasks. It will do so in a variety of ways, using a plethora of

techniques. Hard copy text will surely remain a primary source of information and ideas.

Traditional critics of information technology have ignored the fact that reading is itself a highly private act. One hopes that as a consequence of reading good literature or listening to good music, certain public activities and manifestations will occur. But whether civic virtue results from high culture has always remained problematical. The act of reading or viewing is still largely a private phenomenon. The forms of such enjoyment will undoubtedly change under the impact of the new information technology. But this does not necessarily mean that changes will be toward either greater free expression or greater totalitarian control. There is only hope that the public and private roles may be meshed, and the thin line between entertainment and instruction will be further reduced, for example, through the mechanisms of videogames performing mediating roles between science and entertainment. This is just a vague hope, however.

Less speculative is that the new information technology represents not a diminution, but an addition to what now exists in the way of publishing potential. It provides a definite value-added attribute to the reading of books in hard copy form. The fears that self-publishing, in effect vanity publishing, will conspire to destroy the foundations of scholarship, fail to reckon with the accepted gatekeeper functions assigned to publishers in modern societies.¹⁰ Just as commercial publishing has flourished in an era of widespread photocopiers, it can also be expected to survive in an age of electronic dissemination of information.

The conversion of the television screen into a key feature in home and office information dissemination has permitted new links between instructions and enjoyment in mass culture; everything from television cable to video games, i.e. wideband, while at the other end, there is promulgation and

promotion of scholarship and research in electronic media forms, i.e. narrow-band. At the moment, it is the quantity of new forms rather than their quality that is most obvious.

It is hard to predict just how people will apportion their time in this new world. There is also the problem of how many will be excluded. The new information technology provides so many options that a new paradigm will evolve, in terms of how an individual makes decisions about apportioning his time and energy, rather than how his interests are narrowed down. It might be the case that democracy will dissolve under the weight of excess choice and decision. In this scenario, there will be a Balkanization rather than a pluralization effect. A less theatrical view of this new pluralism is the emergence and evolution of a knowledge industry infrastructure which solves the problem of access and delivery of information and the ideas it has created, but at a much higher delivery rate and level of abstraction than previously possible.

POLITICAL EFFECTS AND CONSTRAINTS

The political problem during the final years of the Twentieth Century is much less the amount of scientific information and material available than the integration and accessibility of the value of that information. Such a plethora of technical options has been made available that society is threatening to fall behind in the orderly processing of data and information as such. Democracy works best within a set of commonly accepted guidelines and acted-upon procedures. It is not reducible to pure choice within the confines of a normless external environment. If the potential for a near-insoluble problem does exist within the new technology, it is less with its totalitarian capabilities than its anarchic consequences. The multiplication of options may well increase the chances of anomie and normlessness, rather than in a purely narcotizing dysfunction based upon an information overload.

There is no common computer protocol or language that is able to synthesize the roughly 200-300 databases currently available. The U.S. Department of Defense's promulgation of ADA as its routine language is an attempt to impose order within the information resources of the armed services. But so far, this has added only one more systems capability for users. There are other problems of anarchy; for example, the definition of what constitutes bona fide research results, when systems such as COMTEX are available in which one can publish work in progress. Problems without precedent are beginning to emerge in evaluating quality; refereeing techniques and gatekeeping groups, for example, each of which formerly decided about the quality of performance or of information. Academic gatekeepers may not be adequate for screening such a superabundance of data. These are only some of the enormous problems that we are going to be confronted by in the very near future. They are not simply, or even primarily, limited to blockage and prevention of access, but rather to the dilemma of what happens when access becomes more important to a society than ownership, and recompense for producing information increasingly becomes the author's responsibility. If proprietary considerations yield to considerations of availability and access to information, issues of a new sort emerge for Western societies as a whole.

If researchers have had a difficult time providing a conceptual map of the relationship between the new information technology and political democracy, it is not for lack of data, or because much attention has been focused upon the short-run consequences for publishers of hard copy, but because there remains a deep suspicion that technology in and of itself creates unanticipated negative effects. Theories of a new information technology have highlighted a generation gap, between an older generation for which hard copy is a symbol of status, achievement and arrival, and a young generation of people for whom the video cassette, videotex and video games have become paramount

forms of conveying information. The heroes of the young are not the authors of novels, but figures who dominate television, whether they be in news-casting or commercial entertainment fields.

Many of the presumed weaknesses of the new information technology concern assumptions about the appropriateness of certain formats to deliver data and ideas to precise audiences. In fact, the multitiered approach can adjust for such shortcomings. For example, the video keyboard may be a dominant force in matters of cultural transition and transformation, but when people go to an airport they may prefer to buy a paperback book rather than carry a video-pack in their knapsack. Even with minituarization, electronic nonprint media may not be preferred for delivering information, certainly not in the foreseeable future. Data systems of the future will be multitiered, and the costs involved will have to be relative to the benefits expected. In the 1980s, it is still cheaper and far easier to put paperback volumes into an airport newsstand than to put videopacks into the same space. But even at such an obvious level, researchers should be cautious. Western societies are moving towards a high degree of miniaturization, as a result of microchips which are so high-powered that it may soon be possible to consider small, portable, personal screens that are capable of providing the kind of information currently carried chiefly by hard copy text. The portable television already provides a model for such new delivery processes.

The relationship between political democracy and the new information technology is by no means either uniform or mechanistic. In rapidly developing areas, nationalism may conflict with the wide use of a technology that has foreign or colonial origins. Antidemocratic constraints come masked in hostility to foreign ideas, influences, artifacts and scientific systems. The intense desire for national autonomy not infrequently spills over into tightly knit controls over the new information technology. As one observer

has shrewdly noted about Brazil (a prototypical country in this contradiction between national interests and democratic aims), there are definite limits to such restraints.

On the one hand, there is tremendous demand from all sectors of the economy for computer hardware and software; on the other, there is the principle of diminishing the scientific, technological and economic dependence on the country. Controls on information flow can be justified on the basis of national security, privacy, economics or nationalism. The restrictions, however, are often a two-edged sword: they may protect the country in one way and injure it in another.¹¹

In other words, the need for socioeconomic developments involves maximum participation in the international exchange of information and ideas, even as the need to protect national interests may seek elites to limit such maximal use.

Resources restraining the further distribution and disbursement of information are often viewed as negative, i.e. nationalist regimes, military cliques, political elites. But probably the greatest single impediment to the massification process is the influence of scientific and professional societies representing groups who believe that the intellectual property belongs to creators and authors. Without once again examining the merits and demerits of the debate on the right of public access versus the obligation to recognize proprietary claims, it should be appreciated that the notion of confidentiality, when carried to its ultimate logical conclusion, is also a powerful restraint on the democratization of culture.¹² Everything from the statutory protection of privileged data to the development of limited partnership arrangements between universities and industries has had a dampening impact on the broad use of information and ideas. Dorothy Nelkin¹³ has expressed the current contradictory tendencies within professional life with great persuasion. She indicates that the response of scientists

rests on a notion that scientists have a "right" to control their research, that autonomy is necessary to maintain integrity, to avert the misinterpretation of premature data and to protect their "stock in trade." Those who request data claim the "right to know" as an essential condition of democracy. Government agencies claim the right to information as part of their obligation to ensure responsible use of federal funds, to meet policy goals or to maintain national security or law enforcement in the public interest. Contradictions persist, reflecting the deep ambivalence within science about its cognitive and practical dimensions. Is science the pursuit of truth or useful knowledge, a carefully disciplined process or a professional instrumental activity? The ambivalence so apparent in the disputes over the control of research suggests that there have been significant changes in the social role of science and in the importance of research.

These disputes are part of a larger struggle to renegotiate relationships between science and the public that were established at a time when science was a very different social enterprise. This suggests that the new information technology enlarges the possibilities of greater access and hence the wider practice of democracy, and produces greater limitations to new data and information and hence narrower opportunities for the practice of democracy. The new technology heightens the awareness of democracy as a problem, but does not in or of itself lead to either democratic or dictatorial outcomes.

Arguments over the present-day feasibility of the electronic or nonprint journal often center on whether "garbage" will be displayed on the video terminals: articles which are not refereed, properly edited or scientifically worthwhile. In point of fact, most scholars do not share the publishers' concerns. Only 20%, by one estimate, view this as an issue. Moreover, the arguments tend to be prima facie spurious: first, refereeing a journal in electronic format is no more or less difficult than for a standard journal. Indeed, the refereed report can much more easily become part of the article being viewed than the article being read. Second, the quality of editing varies so much in printed journals, that one can scarcely expect a worse situation to occur in the videotex formats. Finally, the quality of scientific research is not so much the issue as the early establishment of

primacy in the discovery process. In this sense, the new information technology can only speed up the transmission of ideas.¹⁴

What is likely to take place, in the short run at least, is the use of microprocessor-based technology as a prelude to hard copy format. There are still some clumsy aspects of electronic publishing: unavailability of terminals and unreliability of computers. But again, these are short-term considerations. In the larger sense, or at least in the political sense, the new developments clearly change the nature of scholarly procedures. But just as clearly, they do not alter the demands for quality control of research. To be sure, the sophistication of information technology provides unheard-of checks upon fraudulent reportings and preliminary findings published in multiple sources, that simply have not been available in the past. Through the active involvement of the author in copyright reassignments, the new media provide more in the way of service to the scholarly community than threats to that community. The burden of this position is neither to frighten those involved in traditional publication nor necessarily to celebrate the new technology. The plethora of new developments in information dissemination provides a potential for democratic expression previously unavailable.

CONCLUSIONS

Hopefully, this paper has indicated that problems created by the new information technology are not purely pedestrian, but is rather a struggle between economic and educational classes, i.e. between those who do and those who do not have the ability to manage, manipulate and massage the new information technology for their own goals. We may well reach a point where a sophisticated portion of the population becomes capable of managing information technology and participating at the highest reaches of democratic participation, while another stratum of people is created, for whom the new information technology remains an eternal mystery, and who are largely

unable to master the trends and tendencies of the near future. Whether this will be played out in yet another expression of class struggle, or becomes a new way in which society is benignly stratified remains to be seen. But it should be evident that the uneven distribution of hardware and software components makes for a special problem in democracy; a more costly one than the uneven distribution of books.

The issue of computer literacy is nothing more than the issue of literacy and its uses, written largely for a future that has arrived. Whatever its stage of evolution, technology neither automatically opens or closes possibilities.¹⁵ It does provide for new options, and hence new dangers and opportunities. The new information technology contains potential for self-correction, including even larger numbers of people who at present are profoundly blocked from the mainstream of democratic participation. These remain political no less than technical challenges. Hopes for future developments in the area of information technology require a sense of user needs even more than manufacturer capabilities. We are simply once more recreating a post-industrial environment in which life and death issues are replicated on a canvas; this time called information rather than environment.

The central problem posed by the new information technology is the possibility of reaching, even more profoundly, those capabilities of utilizing or accessing available, cumulative knowledge, and reducing the numbers for whom knowledge remains a mystery. The problem of illiteracy now extends beyond words into electronics. The task of democracy is not simply to widen the horizons and opportunities created by information technology, but more significantly, to increase the number of people involved greatly. Democracy is not simply a question of options and choices for elites, but a social responsibility of the widest number of people for their inclusion in the decision-making framework. At this level, the problems caused by the new information

technology have not begun to receive proper attention; yet it is at this level of mass participation and technical literacy that the issue of democracy in the next century will come up against its heaviest challenges, and hopefully, will provide its most innovative responses.

REFERENCES

1. J. Scholl, "The Videotex Revolution," Barron's National Business Finance Weekly. Vol. 62, No. 3 (2 August 1982), pp. 608.
2. P. J. Vinken, Information Economy, Government and Society. European Information Providers Association, United Kingdom (1982).
3. N. Garnham, "The Information Society is also a Class Society," in Proceedings of the EEC Conference on Information Society, ed. U. Barry. National Board for Science and Technology, Ireland (1982).
4. D. Lacey, "Publishing and the New Technology," Books, Libraries and Electronics: Essays on the Future of Written Communication. Knowledge Industry, United States (1982), p. 85.
5. B. Berelson and M. Janowitz, (eds.) Reader in Public Opinion and Communication. Glencoe: Free Press, 1953, pp. 309-16; 469-82.
6. G. B. deHuszar (ed.), The Intellectuals: A Controversial Portrait. Glencoe: Free Press, 1960, pp. 437-552.
7. B. Rosenberg and D. M. White (eds.), Mass Culture Revisited. New York: Van Nostrand Reinhold, 1971, pp. 3-21.
8. J. Sendres, "The Electronic Journal of the Future," American Sociology, Vol. 11, No. 2 (1976), pp. 160-64.
9. F. W. Lancaster, Libraries and Librarians in an Age of Electronics, "Information Resources" (1982), pp. 199-203.
10. L. A. Coser, C. Kadushin, and W. W. Powell, Books: The Culture and Commerce of Publishing. Basic Books, USA (1982); also M. Stanley, "Marketing as a Liberal Art," Scholarly Publishing, Vol. 13, No. 4 (July 1982), pp. 293-308.
11. V. Rosenberg, "Information Policies of Developing Countries: The Case of Brazil," Journal of American Society Information Science. Vol. 33, No. 4 (July 1982), pp. 203-07.
12. I. L. Horowitz, "Corporate Ghosts in the Photocopying Machine," Scholarly Publishing, Vol. 12, No. 3 (July 1981), pp. 299-304; also I. L. Horowitz and M. E. Curtis, "The Impact of Technology on Scholarly Publishing," Scholarly Publishing, Vol. 13, No. 3 (April 1982), pp. 211-28.
13. D. Nelkin, "Intellectual Property: The Control of Scientific Information," Science, Vol. 216 (May 1982), pp. 702-08.
14. W. J. Broad, "Journals: Fearing the Electronic Future," Science. Vol. 216 (May 1982), pp. 964-68; response by B. Lewin, J. Raben, L. H. Weiler, and C. M. Goldstein, "Electronic Publication," Science, Vol. 217, No. 4559 (August 1982), pp. 482-84.

15. B. D. Ruben, J. R. Holz, and J. K. Hanson, "Communications Systems, Technology and Culture" in H. F. Didsbury, Jr., Communications and the Future. World Future Society, USA (1982), pp. 255-66.

DESIGN OF A SELF-ORGANIZING DOCUMENTATION SYSTEM

G.J. Dalenoort

Institute of Experimental Psychology
University of Groningen
P.O. Box 14
9750 AA HAREN
NETHERLANDS

August 1983

ABSTRACT

On one hand, documents often refer to other documents, written before. On the other hand the same documents are cited by other documents, written later. The reasons for being referred to or being cited are of a relative large variety. Documents that pertain to a certain topic, to a subdiscipline, or to a discipline, thus together constitute an elaborate network. Similarly the concepts described in the documents, constitute a conceptual network, that in fact is a part of the complete network of a natural language. Such networks may be considered as passive data networks, but also as active networks, of which the evolution of the excitation patterns constitute information-processing. On the basis of two assumptions, such networks may be considered as self-organizing, one on control of overall excitations, the other on the mechanism of the coming into existence of new connections.

1. Introduction

Our brains are very well suited for tasks of information-processing that require reasoning by analogy, generalisation in inductive reasoning, recognition of patterns with meaning. But brains are not well suited for tasks involving details of the past, or patterns consisting of large numbers of seemingly random features.

These properties of brains create a vicious circle for the problem of retrieving documents that have been stored in the past, because we have often forgotten details that are essential to identify documents by machines. With traditional algorithmic and serial methods of information storage and retrieval, the possibilities to retrieve information on the basis of inaccurate or incomplete cues, are very restricted. With alternative methods of parallel processing, based on neural networks, and on knowledge from cognitive psychology, it seems possible to break the vicious circle. This becomes plausible by the fact that we, with our brains, may be poor in remembering where we have stored a document, but are very good in recognizing the item sought for, once we have found it. It is precisely this kind of fact that we are trying to model, in theory, and also in computer simulation. We would then construct a computer program according to traditional algorithmic and serial design, but at a higher level of interpretation it simulates a parallel process that may be expected to reflect cognitive processes to larger extent than the models based directly on the computer analogy.

This new approach has the additional advantage that we must explicitly consider the problem to make the system self-organizing, since brains can learn. At the level of neurons, where the actual structural changes must take place, only criteria of physical, anatomical, and physiological nature can play a role, which means that learning must finally be based on mechanisms that are independent of the content of incoming information. Of course, since we know that learning can be meaningful, we must be able to interpret the final accumulated results of many low-level steps in meaningful terms. This approach will also allow us to give an interpretation of the concept of meaning, that avoids the problems that seem to be inevitable within the paradigm of information-processing based on current computers.

Scientific articles and books refer to each other, and thus constitute elaborate networks. The concepts of language also constitute an elaborate network, in linguistics and artificial intelligence generally indicated by the term 'semantic networks'. In traditional approaches both kinds of networks have been treated as

passive networks, on which a separate set of procedures can act. In the present approach, we attempt to construct a network and processes at a lower level of aggregation, in analogy with neural networks, that are to serve as minimal functional models of the brain. Since neurons are active elements, neural networks are also active. Therefore, we shall construct the conceptual and document networks based on them, also as active networks, in contradiction to passive data networks with a set of corresponding procedures. The latter will be incorporated in the structure and dynamics of the active networks.

In this approach it is much more difficult to follow a strictly formal design, as is common in artificial intelligence, and related fields. On the contrary, it should be especially useful in domains of fuzzy knowledge and reasoning, where formal approaches are until now far too restricted to be useful. As a consequence, systems to be constructed on the basis of this approach, are likely to be useful in the first place as personal documentation systems, and less as general systems.

2. Clustering in networks

Scientific articles are not isolated. On the one hand they refer to articles written earlier, on the other hand they are cited by articles written later. The articles in a field of science thus constitute a network that however is very non-uniform. Clusters occur for several reasons. Authors will predominantly refer to articles they know well, and concepts are not uniformly related to each other, they also cluster into themes.

This phenomenon of clustering is not restricted to documents. We can also observe it in social groups, and it is a basic postulate of the so-called theory of cell assemblies in neural networks (D.O. Hebb, 1949, see also Dalenoort, 1982). In fact, clustering is a very general phenomenon. Hebb's most important contribution to the theory of neural networks is the indication of a mechanism that was already known, underlying the formation of cell assemblies. The mechanism was already formulated by Tanzi (1893) in relation to learning, when the existence of synapses, the structures connecting neurons, were not yet known. The mechanism can be stated as follows. When two neurons for some reason are active (fire) simultaneously or almost so, existing connections will strengthen, or new connections will be formed. The direction of the connection will of course correspond to the order in time of the events leading to the activities of the neurons.

This mechanism, henceforth to be called Tanzi-Hebb mechanism, can in fact be formulated such that it has validity in general. Social groups will also emerge

on the basis of initially random encounters, random from the point of view of the individuals involved. Similarly, concepts are related on the basis of the structure and order of the world in which we live, and will cluster on the basis of non-random occurrences of events. Therefore, we can formulate a general mechanism:

Simultaneous, or nearly simultaneous activity in two members of a set of systems, results in the beginning or strengthening of interaction between the two systems, with the result that activity in one will enhance, or weaken, activity in the other. The direction of interaction may be one way or both ways.

On general grounds, we must assume that the growth of such interactions will saturate. No other mechanism of equal explanatory value has been formulated, apart from variations. Moreover the principle provides immediate insight into the process of the emergence of social groups, as well as in the process of the development of concepts, in the individual as well as in scientific fields. Further it is so basic that it seems unlikely that it can be reduced to still more basic principles; of course it builds upon primitive concepts like 'object' and 'interaction'.

The Tanzi-Hebb mechanism (or principle, depending on the context) obtains much deeper significance, especially with respect to the representation of knowledge, if it is combined with the conception of complexity, as described by H.A. Simon (1962). He argued that complex systems can only come into existence on the basis of the existence of levels of stability, at each of which modules can survive long enough to serve as building stones for the construction (guided) or emergence (self-organization) of modules at the next, higher level of complexity. The need for such an organization becomes especially compelling with the consideration that in self-organizing processes the large majority of events leading to more complex structures are binary encounters, possibly enhanced by catalysts that can make such encounters more favourable. It seems intuitively clear that the bonds between modules at lower levels of stability (not to be confused with lower stability!) will be stronger than those between more complex modules (for example: the forces between particles in atomic nuclei are much stronger than those between the surrounding electrons and the nucleus, or those between atoms). Simon expressed this by means of the term 'near-separability': in splitting a module with relatively little energy, it will in the first place separate such that smaller modules (of a lower level) will remain intact. Larger modules are near-separable with respect to smaller modules.

These principles of complex organization and structure, and of the way they come into existence, also apply to the set of articles of a field of science. In the next sections we shall consider how a documentation system might be designed on the basis of these principles.

3. Cell assemblies as excitation patterns

Once we have noticed a relation between two articles, it is likely that in our mental representation of the articles, an interaction will start to grow in strength. The cause for that growth must be a common topic, or an analogy. An interaction between scientific articles can thus only occur within our mental representations. Stating that scientific articles interact, is only a short way of stating that some people relate them. In the case such an interaction is not (yet) a general state of affairs for all scientists concerned, it is a "personal affair". To that extent our considerations lead to the design of a "personal documentation system".

Two important observations can be made here. In the first place the question arises to what extent it is an illusion that "objective" documentation systems are sensible, for general use? The wide variety between persons in their appreciation of articles, at least poses the problem whether the common cores of some fields of science are large enough to justify a common data base. The experiences of many users of such systems may shed light on this problem.

Secondly, the fact that interactions between scientific articles can only occur within the mental representations of scientists, points to an important link between networks of articles and cognitive networks. We shall here not explore the consequences of that link in a systematic fashion.

Most current documentation systems consist of a passive data network, from which items can be retrieved on the basis of their properties. Such properties may be stored with the item itself, they may be present on lists of specific properties, referring to corresponding documents, or documents may be reached via other documents, in part by way of inference. All of these methods are usually implemented in terms of a separate set of procedures, that correspond in detail to the organization of the passive data network. Both are programmed on the basis of a specific design. Once a system has been constructed, it is very difficult to make alterations that do not fit the original design. Moreover, it is extremely unlikely that we could find ways to construct such a system by way of self-organization, i.e. that it would change its own structure on the basis of new information, similar to the human brain that learns. We shall

now discuss a design for a network of documents that in principle seems more like the network of human knowledge.

Instead of representing documents as nodes in a passive data network, on which separate procedures can act, we shall consider documents as active nodes that can excite and inhibit each other. That makes the distinction between data and procedures superfluous: the development of excitation patterns in the active network, is to be equivalent to the action of procedures on the passive network. As stated before, I have constructed the design of such systems from the study of the ways neural networks could process information. Neurons can in first approach be considered as simple threshold elements. If the excitation level rises above the threshold, the neuron fires and subsequently contributes to the excitation levels of the neurons to which it is connected, positively or negatively. There is a continuous decay of the excitation level, so that incoming excitations must occur close enough in time to make a neuron fire.

Single neurons seem too simple elements to serve as the carriers of concepts in the brain. Moreover the memory traces then would be too vulnerable. Therefore the hypothesis was made that groups of neurons serve as the material basis for memory traces. Further study has made it clear that such groups cannot be simple groups, the membership function must depend on the current context. This has led to the idea that the material basis of concepts, in the widest sense, including actions as well, must be excitation patterns. The cell assembly must then not anymore be considered as a group of neurons, but as a dynamic entity, to which individual neurons contribute in varying degrees, depending on current context. This is not to imply that the excitation patterns can exist independently of specific neurons, that the neural network could be considered as a medium for excitations.

Other considerations, on the basis of the restricted capacity of human short-term memory, have led to the postulate that the excitation patterns must have characteristics of a particular shape.

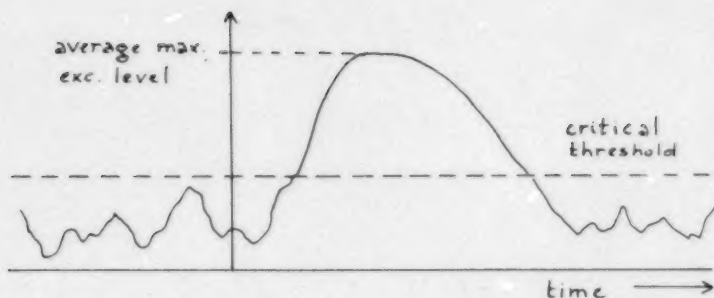


Fig. 1. Excitation characteristic of a cell assembly

Excitations under the critical threshold can be ongoing at the same level, but once they reach that threshold, they grow fast and autonomously to the maximum level, to decay again afterwards. It is reasonable that the average overall excitation of the network is controlled, and remains more or less constant. If this would apply only to the above-critical-threshold activity, and if the maximum values of the excitation patterns are equal, at every time only a restricted number of excitation patterns can be simultaneously active.

It is not difficult to construct networks with such behaviour, but we do not yet know which structural parameters will ensure that they come into existence by self-organization, which must be effected by the material carriers at the lowest level, the neurons, according to the Tanzi-Hebb mechanism, as described in Section 2.

Thus there is a conceptual difference between a neuron as a threshold element and a cell assembly or excitation pattern. The former only contributes to the excitation levels of other neurons when it fires, whereas the latter contributes to other excitation levels in proportion to its own excitation level, which must be interpreted as the fraction of all neurons firing, or as the properties of the maximum excitation level of the whole pattern.

On the basis of these considerations for neural networks, we shall represent documents as nodes in a network, with a critical threshold, and an excitation characteristic like depicted in Figure 1. The precise value of that threshold

and other parameters remain to be determined from simulation studies. The emergence of new links, or strengthening of existing ones, must be extrapolated from the Tanzi-Hebb rule for individual neurons in the case of the neural network.

In line with the considerations before, we must also allow that documents will cluster, in a modular hierarchy, corresponding to the way we classify the corresponding scientific discipline in subdisciplines, topics, etc. We must expect, as stated before, that such a modularly hierarchical organization automatically emerges in a self-organizational process, by consequent application of the Tanzi-Hebb rule.

Since we have restricted this paper to considerations of design, we shall not elaborate on these topics. But it may have emanated that this approach is rich in potential, although it may be not easy to exploit.

4. Networks of documents

In order to become more specific I shall specify the nature of a document, as a representant of a scientific article. It would be too uneconomic to include complete articles. The identifier of a document in the system is the name of its first author, and the year of publication, possibly with an additional character for documents that would otherwise have the same name, e.g. CHARLES-1977a and Charles-1977b.

The contents of a document could be:

AUTHORS

TITLE

PLACE OF PUBLICATION

KEYWORDS

DATE OF DOCUMENT

CHARACTERIZATION (theoretical, experimental, application of a theory, etc.)

EVALUATION (with respect to the aims of the owner, or team using the system)

REFERENCES (only the important ones, and leaving out the ones mentioned casually, but possibly including documents that have not been mentioned)

CITATIONS (names of later documents mentioning this one).

The references and citations can be of various sorts. They are the expression of direct interactions between scientific articles. The most important are (without an order of importance being implied):

- elaboration of a theory
- application of a theory
- experimental confirmation of a theory
- experimental refutation of a theory
- integration of a theory into a more general one
- construction of a new theory from an experimental finding
- formalization of a heuristic theory
- correction of a theory or experiment
- incorporation into a review.

In order that the specifications of such links obtain a representation in the system, we must also introduce nodes for these meanings. The fact that article B is an elaboration of a theory described in article A, is then represented by the excitation of the node for the document B (abstracting article B) by the combined activity of the node for document A, and the node for "elaboration of a theory". It is important to realize that a node cannot send excitation to a single specific other node. In a sense it is blind, and it sends excitation along all outgoing connections. In the present case the two sets of outgoing excitation fans coincide at the node for document B, which will thus be the first to reach its critical threshold, and to get excited to its maximal value. To make such combined activity of two nodes more specific, it seems advisable to introduce a subnode, as indicated in Figure 2.

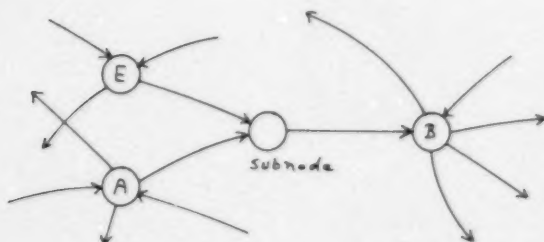


Figure 2. The combined activities of node E, for "elaboration of a theory", and of node A, excite a subnode of node B, which then in turn is excited

The values of the connection strengths and critical thresholds are to be determined, to obtain appropriate functioning. It remains a matter of taste whether one wants to count the subnodes of a node, as in Figure 2, as to belong to node B or not. Nodes would then become complex entities.

An important observation must be made as to the statement that node E in Figure 2 represents "the meaning of 'elaboration of a theory'". This node only obtains this meaning by way of its specific contribution to the link from A to B. The label put on node E has no meaning for the system itself, but only for our description of the system. In the actual system, we could only retrieve the, say, written representation of that meaning, by activating the node, such that it produces output on a typewriter or lineprinter. Similarly, a document node only obtains meaning from its interactions with other nodes. Meaning thus is purely relationally defined. The same applies to any linguistic or artificial intelligence system, and this observation can solve many epistemological riddles and paradoxes, that can be found in the literature ("the meaning of meaning"), In other articles I have discussed this issue more elaborately, also in relation to primitive concepts (Dalenoot, 1983).

We already have extended the document network with conceptual nodes, to specify links between nodes. In the next section we shall elaborate on the conceptual network in full. Connections between nodes can have different strengths. We can let these strengths change (increase or decrease) when documents happen to be simultaneously active.

It is of course in principle also possible to use the network in traditional manners, e.g. for a search on keywords, possibly weighted by numbers corresponding to the evaluation. But the most interesting use seems to be search on the basis of fuzzy and incomplete information, such as that the document was entered about two years ago. This must lead to priming all nodes of two years ago with some amount, and somewhat weaker those between one-and-a-half and three years ago. We may even add information on how the reference to the document was obtained, because this often is the sort of information we remember. Of course, all such information can also be incorporated into traditional systems. The advantage of the present approach is that the documents may be of different formats. Moreover, operation of the system does allow a certain amount of inconsistency, as long as reaching the threshold of one node is only weakly or not dependent on contradictory information in a relatively weakly related part of the network. People can also live with many inconsistencies and irrationalities.

All of these ideas must of course be tested in real-life situations and continued use of the system.

5. The conceptual network and conceptual processing

We have seen that for the construction of the document network we had to introduce nodes that specified the nature of the relations between documents. Since we are not interested in the retrieval of documents in itself, but in conceptual analysis on the basis of documents, we may as well try to add information to the network, in relation to the stored documents, that can help to retrieve the appropriate documents. We can augment the document network with a conceptual network, that provides the possibility to specify documents searched for, or make requests, in a more sophisticated manner than by boolean combinations of keywords alone, and also in an implicit manner. The system may then come to explicit specifications on the basis of fuzzy associations among excitation patterns and chains of associations (cf. the phase sequences of Hebb, 1949).

We shall call these networks ACNs, for active cognitive (or conceptual) networks, in contradiction to the usual passive semantic networks, that serve as data for procedures. The so called spreading-activation networks of Quillian (1969, but already originating in 1962), and their elaboration by Collins & Loftus (1975), are intermediate between both. These systems operate by means of tags that are put on nodes to which the nodes are connected from which a search process starts. If tags from different nodes coincide, semantic and syntactic tests are carried out to see if the coincidence is useful. In a sense the system is binary, whereas in our system excitation levels may be built up from more than two nodes, although we may expect that the majority of "critical-threshold passers" will originate from the combined activity of only two nodes. But it is precisely the few that are due to more complex interactions, that may produce interesting results.

In Quillian's system the tests are carried out by separate procedures. In our system we want to incorporate these into the network. Other important differences are the decay of excitation levels in our system, and self-adaptation of the structure of the system on the basis of ongoing activity (self-organization). We also note that there is no central controlling agent, all control as to content is collective. There only must be a mechanism for the control of average overall activity.

In the first versions of the system we shall restrict ourselves to a very limited set of fixed-format requests to, and answers from the system. The slots of the formats can then be filled on the basis of the excitation patterns, as developing in time. Since the kind of network described is a general purpose representation, it is in principle possible to incorporate syntactic rules into the network. It remains to be investigated whether it is a manageable representation, that can be easily related to our traditional ways of thinking. The relation between both kinds of representation can be described as follows.

In the network, an enormous amount of activity is going on at any time. Only a restricted number of excitation patterns can simultaneously be at their maximum level, due to the limited average overall activity, and the autonomous rising of activity above the critical threshold. If one excitation pattern, together with incoming activity, excites another excitation pattern, the former one, or a third one, must be suppressed. The relation between both representations is thus that the temporal sequence of maximum excitations in the network, corresponds to the algorithmic sequence of the traditional serial systems. The hope is that the design rules we put into the algorithms, on the basis of our high-level knowledge, can be replaced by the massive low-level processes in the network. The high-level excitations emerge out of the competition of the low-level processes, under influence of inputs. By means of the Tanzi-Hebb mechanism the system adapts its structure, so that it learns.

An important aspect has not yet been considered in sufficient detail, the modular hierarchy. When nodes often are active together they will give rise to the formation of a new node. In neural networks this can be understood in terms of the underlying neural processes. Initially both parent nodes will become excited to their maximum, but once the new node is sufficiently strong (i.e. when its incoming excitations are strong), it must suppress the activity of its parent nodes. Finally, the combined, and synchronized activity, below their critical thresholds, of two or more nodes must be sufficient to raise one node above its critical threshold. That node then represents the integrated concept, a chunk in the terms of cognitive psychology. An example is a word that consists of letters, that in turn consist of features (cf. McClelland & Rumelhart, 1981).

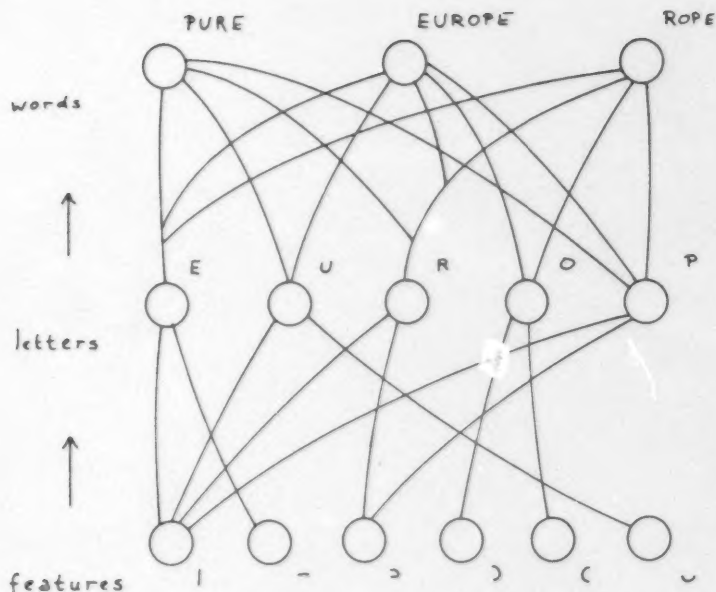


Fig. 3. Incomplete design of a network for features, letters, and words. It is obvious that subnodes (or interactions between connections) are necessary to account for letter order, and spatial organization of features. (cf. McClelland & Rumelhart, 1981).

Not all concepts can easily, or even at all, be integrated into new concepts. This may be a consequence of training. For example, letters can easily be integrated into new words, but mixtures of letters (a, b, c, ...) and figures (1, 2, 3, ...) not. This is probably due to the basic differences of the classes of letters and of numbers, including aspects of pronunciation. We must conclude

that nodes that (already) belong to fundamentally different conceptual structures, cannot be so easily integrated. To some extent integration is possible, for example E1 has definite meaning for European chess-players and European car-drivers.

At present, the topic of controlled (conscious) and automatic (semi-conscious or sub-conscious) processing is important (Schneider & Shiffrin, 1977). The distinction is of direct relevance for the possibility of formation of integrated concepts (chunks).

Let us finally give the outline of one example of how a network system might handle a specific request. It would lead too far, and require too much speculation to specify a complete analysis at this stage of research.

Consider the request:

"Has Jones written anything on document retrieval?"

The problem is to find combinations of two or more excitations that can excite further nodes that can lead to an answer. It is plausible that the specific word order, together with the question mark, sets the system to an answering mode (in contrast to a message for the system). We assume that this setting consists of the priming of certain nodes corresponding to syntactic rules, and superpositions of rules, that together with semantic nodes can produce output. In Figure 5 a simplified scheme is presented, where the syntactic nodes are indicated by A1, A2, ...

The words of the question will excite word nodes, which in humans would proceed through auditory or visual channels. The excitation of a word node in itself is not yet sufficient for understanding, only further interactions can produce 'meaning' (cf. Section 4).

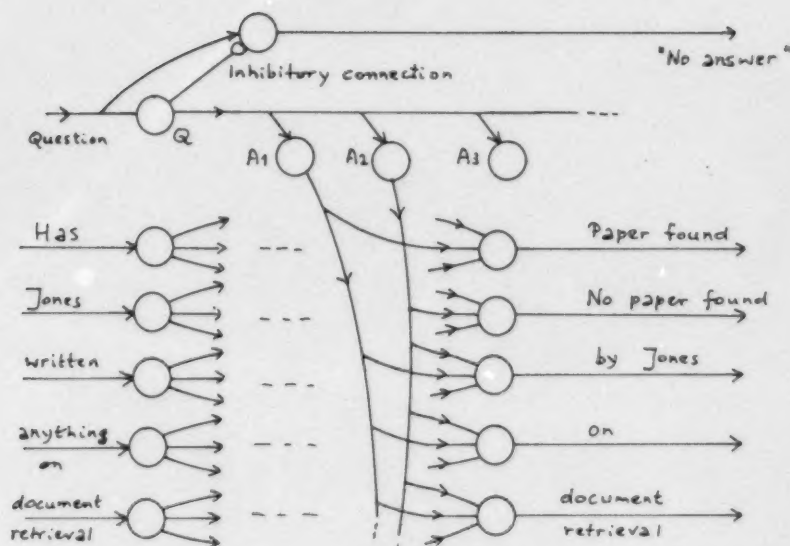


Fig. 4. Scheme of the production of an answer to the question "Has Jones written anything on document retrieval?" Q stands for "question", A1, A2, ... for possible answers, or syntactic structures that can combine to answers.

The order of the words in sentences must be represented by further interactions between nodes, or by means of additional intermediary nodes. For example, to fill in "Paper by on" The networks designed by Grossberg (1978) for the explanation of the serial-position effect show how this might be done. (The serial-position effect is the phenomenon that items at the beginning of a learned list, and those at the end, are better remembered than items in the middle, the former because they have had more time to be entered in long-term memory, the latter because they are still in short-term memory. In our system, long-term memory corresponds to structure, and short-term memory to current above-critical-threshold excitations).

This example shows the complexity of the problem. Fortunately, there are two approaches to this kind of problem that can help to solve it. The first is the work by Quillian (1969) and by Collins and Loftus (1975), already mentioned. The second is the work by Wilks (1975) on a system called preferential semantics. In the latter, one of the central concepts is 'semantic density'. Analysis of sentences is based on occurrences of the greatest semantic density. It is performed by means of fixed formats (templates) and procedures. This system is thus not yet based on a uniform representation as attempted here. (A somewhat more elaborate description of the relation of both systems is in preparation, Dalenoort, 1983).

References

- Collins, A.M. & Loftus, E.F. (1975), A spreading-activation theory of semantic processing, Psychol. Rev. 82 (1975) 407-428.
- Dalenoort, G.J. (1982), In search of the conditions for the genesis of cell assemblies: a study in self-organization, J. Social Biol. Struct. 5 (1982) 161-187.
- Dalenoort, G.J. (1983), A theory of neural processes underlying language-processing. (in preparation).
- Grossberg, S. (1978), A theory of human memory: self-organization and performance of sensory-motor codes, maps and plans, in Progress in Theoretical Biology, vol. 5 (Eds. R. Rosen & F. Snell), Academic Press, New York, 1978.
- Hebb, D.O. (1949), The organization of behavior, Wiley, New York, 1949.
- McClelland, J.L. & Rumelhart, D.E. (1981), An interactive activation model of context effects in letter perception: Part 1, an account of basic findings, Psychol. Rev. 88 (1981) 375-407.
- Quillian, M.R. (1969), The Teachable Language Comprehender: a simulation program and theory of language, Comm. ACM 12 (1969) 459-476.
- Schneider, W. & Shiffrin, R.M. (1977), Controlled and automatic human information processing: I. detection, search, and attention, Psychol. Rev. 84 (1977) 1-66.
- Simon, H. (1962), The architecture of complexity, Proc. Amer. Philosoph. Soc. (1962) 467-482. (Repr. in: H. Simon, Sciences of the Artificial, M.I.T.-Press, Cambridge, Mass. 1969).
- Tanzi, E. (1893), I fatti e le induzioni nell'odierna istologia del sistema nervosa, Rivista Sperimentale di Freniatria 19 (1893) 419-472.
- Wilks, Y. (1975), An intelligent analyzer and understander of English, Comm. ACM 18 (1975) 264-274.